

OM of: US-09-540-234-1 to: A\_Geneseq\_0601.\* out\_format : pfs  
Date: Sep 7, 2001 5:05 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09540234/runat\_07092001.145854.19399/app\_query.fasta\_1.507  
-DB-A\_Geneseq\_0601 -QWMT=fastan -SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blowsum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09540234\_EGNI\_1.41 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:

Query: US-09-540-234-1  
Query length: 444  
Database: A\_Geneseq\_0601.\*  
Database sequences: 412676  
Database length: 60623988  
Search time (sec): 55.090000

score\_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
/SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:AAW11704 +		81.50	147.14	0.4108	637
/SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:AAW57224 +		81.50	146.68	0.4165	667
/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW90342 +		77.50	154.42	0.7793	132
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW58634 +		75.50	140.10	1.0782	355
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW58633 -		75.50	139.33	1.86	383
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW58632 -		75.50	138.77	1.89	405
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW58569 -		75.00	139.76	3.68	2385
/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW90342 -		74.50	136.51	2.50	409
/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW90341 -		74.50	134.70	2.64	489
/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW9484 +		73.50	147.09	2.27	116
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW29133 -		73.50	143.97	2.49	158
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW47928 -		73.50	143.97	2.49	158
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW29132 -		73.50	137.90	2.97	288
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW47927 -		73.50	137.90	2.97	288
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW47926 -		73.50	137.86	2.98	289
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW47925 -		73.50	137.86	2.98	289
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW85574 +		73.50	115.61	5.72	2608
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW06844 +		73.00	143.08	2.84	155
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW06843 +		73.00	139.31	3.18	225
/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW01519 -		72.50	119.31	6.36	1462
/SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:AAW97246 +		72.50	108.00	8.86	4472
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW17227 -		71.50	132.93	5.28	307
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW17226 -		71.50	132.90	5.28	308
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW17225 -		71.50	132.83	5.29	310
/SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:AAW38470 -		71.50	107.20	11.23	3910
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAW66462 -		71.50	107.20	11.23	3910
/SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:AAW52971 -		71.50	107.05	11.28	3969
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAW86250 +		70.00	132.66	7.33	229
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAW86251 +		70.00	132.31	7.40	237
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAW86252 +		70.00	132.31	7.40	237
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW79332 -		70.00	125.45	9.05	467
/SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:AAW13319 -		69.50	116.44	13.12	1023
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW75296 +		69.00	135.84	8.26	135
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW75297 +		69.00	135.84	8.26	135
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAW86272 +		69.00	130.50	9.66	229
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAW86262 +		69.00	130.50	9.66	229
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAW86273 +		69.00	130.15	9.76	237
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAW86274 +		69.00	130.15	9.76	237
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAW86263 +		69.00	130.15	9.76	237
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAW86264 +		69.00	130.15	9.76	237

/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW21873 +	69.00	119.43	13.38	684
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW3788 +	68.50	135.07	9.40	131
/SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:AAW56069 -	68.50	126.82	11.98	296
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW76014 -	68.50	126.72	12.01	299
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAW76077 -	68.50	126.72	12.01	299
seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:AAW11704				
seq_documentation_block:				
ID	AAW11704	standard; Protein; 637	AA	
XX	AAW11704;			
AC	AAW11704;			
XX				
DT	02-APR-1997	(first entry)		
XX				
DE	High affinity Na+-dependent, L-Proline transporter.			
XX				
KW	High-affinity sodium dependent L-Proline transporter; rat brain;			
KW	RTN2-2-20; neurotransmitter transporter; ss.			
XX				
OS	Rattus rattus.			
XX				
FH	Key	Location/Qualifiers		
FT	Domain	46..65		
FT	Domain	/note= "Transmembrane domain 1"		
FT	Domain	73..97		
FT	Domain	/note= "Transmembrane domain 2"		
FT	Domain	117..137		
FT	Domain	/note= "Transmembrane domain 3"		
FT	Domain	217..236		
FT	Domain	/note= "Transmembrane domain 4"		
FT	Domain	243..264		
FT	Domain	/note= "Transmembrane domain 5"		
FT	Domain	291..311		
FT	Domain	/note= "Transmembrane domain 6"		
FT	Domain	323..345		
FT	Domain	/note= "Transmembrane domain 7"		
FT	Domain	374..397		
FT	Domain	/note= "Transmembrane domain 8"		
FT	Domain	424..443		
FT	Domain	/note= "Transmembrane domain 9"		
FT	Domain	457..479		
FT	Domain	/note= "Transmembrane domain 10"		
FT	Domain	500..519		
FT	Domain	/note= "Transmembrane domain 11"		
FT	Domain	537..559		
FT	Domain	/note= "Transmembrane domain 12"		
FT	Modified-site	14..15		
FT	Modified-site	/note= "Protein kinase C phosphorylation site"		
FT	Modified-site	41..42		
FT	Modified-site	/note= "cAMP-dependent protein kinase phosphorylation site"		
FT	Modified-site	239..240		
FT	Modified-site	/note= "Protein kinase C phosphorylation site"		
FT	Modified-site	600..601		
FT	Modified-site	/note= "Protein kinase C phosphorylation site"		
FT	Domain	76..97		
FT	Modified-site	/note= "Leucine zipper motif"		
FT	Modified-site	182..183		
FT	Modified-site	/note= "N-linked glycosylation site"		
XX				
XX	US5580775-A.			
PN				
XX				
PD	03-DEC-1996.			
XX				
PF	01-MAY-1992;	92US-0879617.		
PR	01-MAY-1992;	92US-0879617.		
XX	(UYEM-) UNIV EMORY.			
PA				
XX				
PI	Blakely RD, Caron MG, Fremeau RT;			
XX				

BEST AVAILABLE COPY











XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 25-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.



PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 18-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 21-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 21-JUN-1999;	99US-0139899.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140353.	PR 01-SEP-1999;	99US-0151930.
PR 23-JUN-1999;	99US-0140354.	PR 07-SEP-1999;	99US-0152363.
PR 24-JUN-1999;	99US-0140695.	PR 10-SEP-1999;	99US-0153070.
PR 28-JUN-1999;	99US-0140823.	PR 13-SEP-1999;	99US-0153758.
PR 29-JUN-1999;	99US-0140991.	PR 15-SEP-1999;	99US-0154018.
PR 30-JUN-1999;	99US-0141287.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0141842.	PR 20-SEP-1999;	99US-0154779.
PR 01-JUL-1999;	99US-0142154.	PR 22-SEP-1999;	99US-0155139.
PR 02-JUL-1999;	99US-0142055.	PR 23-SEP-1999;	99US-0155486.
PR 06-JUL-1999;	99US-0142390.	PR 24-SEP-1999;	99US-0155659.
PR 08-JUL-1999;	99US-0142803.	PR 28-SEP-1999;	99US-0156458.
PR 09-JUL-1999;	99US-0142920.	PR 29-SEP-1999;	99US-0156596.
PR 12-JUL-1999;	99US-0142977.	PR 04-OCT-1999;	99US-0157117.
PR 13-JUL-1999;	99US-0143542.	PR 05-OCT-1999;	99US-0157753.
PR 14-JUL-1999;	99US-0143624.	PR 06-OCT-1999;	99US-0157865.
PR 15-JUL-1999;	99US-0144005.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144085.	PR 08-OCT-1999;	99US-0158232.
PR 16-JUL-1999;	99US-0144086.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144325.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144333.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159330.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144332.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159638.
PR 20-JUL-1999;	99US-0144884.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0144814.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160767.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160815.
PR 22-JUL-1999;	99US-0145192.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160989.
PR 23-JUL-1999;	99US-0145224.	PR 23-OCT-1999;	99US-0161405.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145918.	PR 26-OCT-1999;	99US-0161359.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161360.
PR 28-JUL-1999;	99US-0145921.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146386.	PR 28-OCT-1999;	99US-0161320.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161992.
PR 03-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147038.	PR 29-OCT-1999;	99US-0162142.
PR 04-AUG-1999;	99US-0147204.		
PR 04-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.		
PR 06-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		
PR 10-AUG-1999;	99US-0148171.		
PR 11-AUG-1999;	99US-0148319.		

alignment\_scores:

Quality: 75.50

Ratio: 1.198

Percent Similarity: 57.273

alignment\_block:

US-09-540-234-1/rev x AAG58632

Length: 110

Gaps: 6

Percent identity: 27.273

PA (JANC ) JANSSEN PHARM NV.

[illegible]

616 .LeuProGlnGlnGlnGlnHisSerHisProAsnThrAlaThrValAlap 632

55 CTTTCCTTCAGATGACCCAC 36

|||||: : : : : |||

632 rPheIleTyrArgAlaHis 638

seq\_name: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW90342

seq\_documentation\_block:

ID AAW90342 standard; protein; 409 AA.

XX AC AAW90342;

XX DT 24-MAY-1999 (first entry)

XX DE G. max truncated SBP2 protein.

XX KW SBP1; sucrose binding protein; SBP2; sucrose uptake; transgenic plant;

XX KW seed; carbohydrate content; soybean.

XX OS Glycine max.

XX PN W09853086-A1.

XX PD 26-NOV-1998.

XX PF 21-MAY-1998; 98WO-US10465.

XX PR 22-MAY-1997; 97US-0047568.

XX PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.

XX PI Chao WS, Grimes HD;

XX DR WPI; 1999-070155/06.

XX PT New modified plant sucrose binding proteins - used to develop  
PT transgenic plants which can have enhanced or decreased sucrose  
PT uptake activity in developing seeds

XX PS Claim 7; Page 39-40; 58pp; English.

XX CC This sequence represents a novel sucrose binding protein, SBP2 isolated  
CC from Glycine max. This protein is used in a method resulting in the  
CC production of a modified plant sucrose binding protein (SBP) which has a  
CC modified amino acid sequence compared to a corresponding wild-type SBP,  
CC and where expression of the modified SBP in a yeast assay system confers  
CC enhanced sucrose uptake compared to the corresponding wild-type SBP.  
CC The products of the invention can be used for producing transgenic plants  
CC which have modified sucrose uptake activity, particularly in developing  
CC seeds. Enhanced sucrose uptake activity in developing seeds may be  
CC desirable where it is an advantage to increase the carbohydrate content  
CC of the seed (e.g. where the seed is the primary plant material harvested,  
CC such as soybean). In contrast, decreased sucrose uptake activity in  
CC seeds might be desirable where the vegetative material of the plant is  
CC harvested. The SBP regulatory regions confer specific or enhanced  
CC expression in developing seeds and so may be used to express any  
CC transgene in developing seeds.

XX SQ Sequence 409 AA;

alignment\_scores:

Quality: 74.50 Length: 168

Ratio: 0.887 Gaps: 8

Percent Similarity: 50.000 Percent Identity: 25.595

alignment\_block:

US-09-540-234-1/rev x AAW90342 ..

Align seg 1/1 to: AAW90342 from: 1 to: 409

437 AGACACGAGTACATGGCTAAGAGTAGTACCACCTACAAAGGGAACGAG 388

43 LysHisGlnCysGlnGlnGlnArgLntyrThrGluSerAspLysArg.. 58  
|||||: : : : : ||| : : : : : |||||  
387 TACTGTAGTACACAGGTGTCCCGACAGCAGCAGCAACACACTATGCTG 338  
||| : : : : : ||| : : : : : ||| : : : : : : : : : :  
59 .ThrCysLeuGlnGlnCysAspSerMetLysGlnGlnArgGluLysGlnV 75  
337 CCAGCGAGTACACGAAGAA..... 318  
:  
75 alGluGluGluThrArgGluLysGluGluHisGlnGlnHisGlu 91  
317 .....GTCTTCGAGTGTGCGACGGTAACAG 292  
|||||: : : : : ||| : : : : : |||  
92 GluGluGluAspGluAsnProTyrValPheGluGluAspLysAspPheSe 108  
291 CACTCAATGGCTCCAGCGGAGTACAGCAAGGAAGTCAATCAAGCGCTAC 242  
|||||: : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
108 rThrArgValGluThrGluGly.GlySerIleArgValLeuLysLysPhe 124  
241 ACAGTAAACAGCACACGAATGGCTCCGATACATCGACGCCATTGGCGTT 192  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
125 ThrGluLysSerLysLeuLeuGlnGlyIleGluAsnPheArgLeuAlaI 141  
191 GTTAGTCTGTACAAACTACGGCTTCAC.....CGACGTC 157  
: : : : : : : : : : : ||| : : : : : ||| : : : : : |||  
141 eLeu.....GluAlaArgAlaHisThrPheValSerProArgHisP 155  
156 TCGAGGGGCATGTTTGTAGACAGCACCTCAAGAGCGGAGGACTGTTTCCA 107  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||  
155 heAspSerGluValValLeuPheAsnIleLysGlyArgAlaValLeuGly 171  
106 TTATGCCGTGCGCAGCTACAGTAGCGGTGTCGGCTTGAGGCTGCTTTGT 57  
||| ||| ||| : : : : : : : : : : : |||||  
172 LeuValArgGluSerGluThrGluLysIleThr.LeuGlu..... 184  
56 CCTTCTCCTCAGAT.....GACCCACTAGGCGCTCGCATACATA 16  
|||||: : : : : ||| : : : : : ||| : : : : : |||||  
185 .....ProGlyAspMetIleHisIleProAlaGlyThrProLeuTyrIle 199

seq\_name: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW90341

seq\_documentation\_block:

ID AAW90341 standard; protein; 489 AA.

XX AC AAW90341;

DT 24-MAY-1999 (first entry)

XX DE G. max SBP2 protein.

XX KW SBP1; sucrose binding protein; SBP2; sucrose uptake; transgenic plant;

XX KW seed; carbohydrate content; soybean.

XX OS Glycine max.

XX PN W09853086-A1.

XX PD 26-NOV-1998.

XX PF 21-MAY-1998; 98WO-US10465.

XX PR 22-MAY-1997; 97US-0047568.

XX PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.

XX PI Chao WS, Grimes HD;

XX DR WPI; 1999-070155/06.

XX PT New modified plant sucrose binding proteins - used to develop  
PT transgenic plants which can have enhanced or decreased sucrose  
PT uptake activity in developing seeds



```
195 GCCAATGGCTGATGTTATCGGACCATTCGTGCTGTGTTACTGTGTA 244
:||||: |||: :|: |||:|||||
39 nProPheSerGlnProGlyAlaProIleLeuCysCys..... 52
245 GGCCTTGATGACTTCCTTGCCTGT.....ACTCCGCTT 276
:|: ||||| |||||
53 .....MetGlyCysCysPheSerArgAlaTyrProThrProLeu 65
:|: ||||| |||||
277 GGAGCCATTGTAGTCTCTTACCGTCGACGACTCGAAGACTTCCTGCT 326
:|: ||||| |||||
66 .....ArgSerLysLysThrMetLeuVa 73
:|: ||||| |||||
327 GTAC....TCGCTGGCAGCATCTCTGTTGCTGCTGCTGGGACACC 373
:|: ||||| |||||
73 IclnLysAsnValThrSerGluSerThrCysCysValAlaLysSerTyrA 90
374 TCTGTACTACACTACTT 390
:|: ||||| |||||
90 snArgValThrValMet 95

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AA229133
seq_documentation_block:
ID_ AAG229133 standard; Protein: 158 AA.
XX AC AAG229133;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 34608.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 16-APR-1999; 99US-0128714.
XX PR 19-APR-1999; 99US-0129845.
XX PR 21-APR-1999; 99US-0130077.
XX PR 23-APR-1999; 99US-0130449.
XX PR 28-APR-1999; 99US-0130510.
XX PR 30-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132407.
XX PR 05-MAY-1999; 99US-0132484.
XX PR 06-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 18-MAY-1999; 99US-0134370.
XX PR 19-MAY-1999; 99US-0134768.
XX PR 20-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
```

```
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
```



```
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148368.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159320.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 73.50      Length: 141
    Ratio: 1.131        Gaps: 6
Percent Similarity: 46.099 Percent Identity: 24.113

alignment_block:
US-09-540-234-1/rev x AAG29133 ..
Align seg 1/1 to: AAG29133 from: 1 to: 158
435 ACACGAGTACATGGCTAAGAGTAGTACCACACAAAGG..... 397
32 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa 48
396 .....GAAACGAAGTACTGTAGTACACAGGTGTCCCCAGACCAGG 357
48 lGlyAsnGlyGluLysLysIleLysSerThrTyrLeuSer..... 61
356 CAGCAAAACAGACTATGCTGCCAGCGAGTACAGCAAGAGTCTTCCAGTG 307
62 .....CysLeuAspAsnLysLysGlnGlnMetGluArgVal 73
306 CTCACGGTAAACAGCAGCACTCAATGGCTCCAAAGCGGAGTACAGCAAGAA 257
74 PheTyrGlyGln.....ArgG1 79
256 GTCATCAAGCGCTACACAGTAAACAGCACAGCAATGGCTCCGATAACATC 207
79 uGlnArgMetAlaThr.....HisGluValValArgSerHisA 92
206 GAGCGCATTTGGCTTTAGTCTTTGACAAACTACGGCTTACCGCTACAGTC 157
92 rArgGlu.....ProGluValArgAspGlyTrpMetGluIle 105
156 TCGAGGGGCGATGTTGAGAGC..... 136
106 GluLeuGlyGluPheGluThrGlySerGlyGluGlyAspAspLysG1 122
135 .ACCCCTCAAAGGGCAGGAGCTGTTTCCATTATGCGTCCGACCTACAG 87
122 uValValMetSerLeuThrGluValLysGlyTyrGlnLeuLysGlyGlyI 139
86 TAGCCGTGTCCGGCTTGGAGGTG 64
139 leAlaIleAspGlyIleGluVal 146

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT: AAG47928

seq_documentation_block:
ID AAG47928 standard; Protein: 158 AA.
XX
AC AAG47928;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60463.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
```

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126564.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140595.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.

```
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 73.50      Length: 141
    Ratio: 1.131      Gaps: 6
    Percent Similarity: 46.099.  Percent Identity: 24.113

alignment_block:
US-09-540-234-1/rev x AAG47928

Align seg 1/1 to: AAG47928 from: 1 to: 158

435 ACACGAGTACATGGCTAAGACGTAGTACCACCTACAAGG..... 397
32 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa 48
396 .....GAAACCAAGTACTGTAGTACACAGGTGTCCCCAGACAGG 357
48 lGlyAsnGlyGluLysLysIleLysSerThrTyrLeuSer..... 61
356 CAGCAAAACAGATATCTGCCAGCGAGTACAGCAAGGAAGTCTTCAGTG 307
62 .....CysLeuAspAsnLysLysGlnGlnMetGluArgVal 73
306 CTCGACGGTAAACAGCAGCTCAATGGCTCCACGGCAGTACACAGGAA 257
74 PheTyrGlyGln.....ArgG1 79
256 GTCATCAAGCGGTACACAGTAAACAGCACACGAATGGCTCCGATACATC 207
79 uGlnArgMetAlaThr.....HisGluValValArgSerHisA 92
206 GAGGCCATTGGCGTTTGTAGTCTGTACAAACTACCGCTTACCCAGCTC 157
92 rgArgGlu.....ProGluValArgAspAspGlyTirpMetGluIle 105
156 TCGAGGGCGCATCTTTCAGACG..... 136
106 GluLeuGlyGluPheGluThrGlySerGlyGluGlyAspAspLysG1 122
```

```
135 .ACCCTCAAAAGGCGAGGACTGTTTTCATTATGCGTCCGCTACAG 87
122 uValValMetSerLeuThrGluValLysGlyTyrGlnLeuLysGlyI 139
86 TAGCCGTGTCGCGCTTGGAGGTG 64
139 leAlaIleAspGlyIleGluVal 146

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AA29132

seq_documentation_block:
ID AAG29132 standard; Protein: 288 AA.
XX AC AAG29132;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 34607.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132407.
XX PR 05-MAY-1999; 99US-0132484.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
```

PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 30-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140595.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 30-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 01-JUL-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 02-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 08-JUL-1999;	99US-0142390.	PR 22-SEP-1999;	99US-0155139.
PR 08-JUL-1999;	99US-0142803.	PR 23-SEP-1999;	99US-0155486.
PR 09-JUL-1999;	99US-0142920.	PR 24-SEP-1999;	99US-0156559.
PR 12-JUL-1999;	99US-0142977.	PR 28-SEP-1999;	99US-0156458.
PR 13-JUL-1999;	99US-0143542.	PR 29-SEP-1999;	99US-0156596.
PR 14-JUL-1999;	99US-0143624.	PR 04-OCT-1999;	99US-0157117.
PR 15-JUL-1999;	99US-0144005.	PR 05-OCT-1999;	99US-0157753.
PR 16-JUL-1999;	99US-0144085.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144086.	PR 07-OCT-1999;	99US-0158029.
PR 19-JUL-1999;	99US-0144325.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144331.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144334.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159329.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159637.
PR 21-JUL-1999;	99US-0144814.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0145086.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160741.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160814.
PR 23-JUL-1999;	99US-0145145.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160981.
PR 26-JUL-1999;	99US-0145276.	PR 22-OCT-1999;	99US-0160989.
PR 27-JUL-1999;	99US-0145313.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145318.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161992.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.		
PR 06-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 08-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		

alignment\_scores:      Quality: 73.50      Length: 141  
Ratio: 1.131      Gaps: 6  
Percent Similarity: 46.099      Percent Identity: 24.113

alignment\_block:

```
US-09-540-234-1/rev x AAG29132 ..
Align seg 1/1 to: AAG29132 from: 1 to: 288

435 ACAGGAGTACATGCTAAGAGTAGTACCCTACCAAGG..... 397
:::|||||:::|||||:::|||||:::
162 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa 178
396 .....GAAACGAGTACTGTACTACACAGGTGTCCCGACAGCAGG 357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
178 lGlyAsnGlyGluLysLysLysLysSerThrTyrLeuSer..... 191
356 CAGCAAAACAGCATATGTCGACGAGTACAGCAAGGAATCTTCGAGTG 307
||| ::::: ::::: |||||
192 .....CysLeuAspAsnLysLysGlnGlnMetGluArgVal 203
306 CTCGACGGTAAACAGCACTCAATGGTCCCAAGCGGAGTACAGCAAGAA 257
:::|||||::: |||||
204 PheTyrGlyGln.....ArgG1 209
256 GTCATCAAGGGCTACACAGTAACAGCACACGAATGGTCCGATAACATC 207
::: ||||| ||||| ::||| ||||
209 uGlnArgMetAlaThr.....HisGluValValArgSerHisA 222
206 GACCCCATTTGGGTTGTAGTCTGTGTACAACACTACGGCTTCACGAGTC 157
-|||::: ::::: ||||| :::::
222 rgArgGlu.....ProGluValArgAspAspGlyTrpMetGluIle 235
156 TCGAGGGGCATGTTTGAGACG..... 136
::: ||| ||||| |||||
236 GluLeuGlyGluPheGluThrGlySerGlyGluGlyAspAspLysG1 252
135 .ACCCTCAAAAGGCGAGGACTGTTTTCATTTATGCGCGTCCAGCTACAG 87
::: ||||| ||||| ||| :::::
252 uValValMetSerLeuThrGluValLysGlyTyrGlnLeuLysGlyI 269
86 TAGCCGTGTCGGCTTGAGGTG 64
:|||||:::|||||:::|||||
269 leAlaIleAspGlyIleGluVal 276

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.AAG47927
seq_documentation_block:
ID AAG47927 standard; Protein; 288 AA.
XX
AC AAG47927;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60462.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
```

```
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 73.50      Length: 141
Ratio: 1.131        Gaps: 6
Percent Similarity: 46.099 Percent Identity: 24.113

alignment_block:
US-09-540-234-1/rev x AAG47927 ..
Align seg 1/1 to: AAG47927 from: 1 to: 288

435 ACACGAGTACATGGCTAAGACTAGTACACTACAAAGG..... 397
:::||||:||||:||||:||||:||||:||||:
162 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa 178
396 .....GAAACGAGTACTGTAGTACACAGGTGTCCTCCACGACGAGG 357
178 lclYAsnGlyGluLysLysIleLysSerThrTyrLeuSer..... 191
356 CAGCAACAGACTATGCTGCCAGCAGTACAGCAAGAAAGTCTTCGAGTG 307
192 .....CysLeuAspAsnLysLysGlnMetGluArgVal 203
306 CTGCACGGTAAACAGCACTCAATGGCTCCAAGCGGAGTACACGAAGAA 257
204 PheTyrGlyGln.....ArgGl 209
256 GTCATCAAGCGCTACACAGTAACAGCACGAATGGCTCCGATAACATC 207
209 uGlnArgMetaIaThr.....HisGluValValArgSerHisA 222
206 GACGCATTGGCGTGTAGTCTTGTACAAACTACGGCTTCCACGACGTC 157
222 lrgArgGlu.....ProGluValValArgAspGlyTirpMetGluile 235
156 TCGAGGGGCATGTTTCAGAGC..... 136
236 GluLeuGlyGluPheGluThrGlySerGlyGluGlyAspAspLysGl 252
135 ..ACCTCAAAAGGCGAGGACGTGTTTCCATTATGCCGTGCCAGCTACAG 87
252 uvalValMetSerLeuThrGluValLysGlyTyrGlnLeuLysGlyGlyI 269
86 TAGCCGTGTCGGCTGGAGGTG 64
269 leAlaIleAspGlyIleGluVal 276

seq_name: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT: AAG29131
```

seq\_documentation\_block:  
ID AAG29131 standard; Protein; 289 AA.  
XX AC AAG29131;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 34606.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

```
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 73.50      Length: 141
    Ratio: 1.131        Gaps: 6
    Percent Similarity: 46.099      Percent Identity: 24.113

alignment_block:
US-09-540-234-1/rev x AAG29131 ..

Align seg 1/1 to: AAG29131 from: 1 to: 289

435 ACACGAGTACATGGCTAGCAAGTACAGCAAGTCTTCGAGTG 397
    ::::|||||:  ::::|||||:
163 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa 179
396 .....GAACGAGTACTAGTACACAGGTGTCCTCCAGCAGG 357
    ||| ||| ||||| |||||
179 lGlyAsnGlyGluLysLysIleLysSerThrTyrLeuSer..... 192
```

```
356 CAGCAACACAGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGTG 307
    ||| ::::  ::::  |||||
193 .....CysLeuaspAsnLysGlnMetGluArgVal 204
306 CTGCACGGTAACACAGCACTCAAAATGGCTCCAAGCGGAGTACAGCAAGAA 257
    ::::|||||:
205 PheTyrGlyGln.....ArgG1 210
256 GTCATCAAGCGCTACACAGTAACAGCACAGCAATGGCTCCGATACATC 207
    :::: ||||| ::::  |||||
210 uGlnArgMetAlaThr.....HisGluValValArgSerHisA 223
206 GAGCGCATTTGGCGTTTGTAGTCTTGTCAAACTACGCGTTTCCACCGACGC 157
    ::::  ::::  ::::  ||||| ::::
223 rGArgGlu.....ProGluValArgAspGlyTrpMetGluIle 236
156 TCGAGGGCATGTTTGTAGACG..... 136
    :::: ||| |||||
237 GluLeuGlyGluPheGluThrGlySerGlyGluGlyAspAspLysG1 253
135 .ACCCTCAAAGGCGAGGACTGTTTTCATTATGCCGTGCCAGCTACAG 87
    ::::  ::::  ||| ::::
253 uValValMetSerLeuThrGluValLysGlyTyrGlnLeuLysGlyGlyI 270
86 TAGCCGTGTCGGCTTGGAGGTG 64
    ::::  ::::  ::::  |||||
270 leAlaIleAspGlyIleGluVal 277

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT: AAG47926
seq_documentation_block:
ID AAG47926 standard; protein; 289 AA.
XX
AC AAG47926;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60461.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD
PF 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
```





```
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
```

```
alignment_scores:
  Quality: 73.50      Length: 141
  Ratio: 1.131       Gaps: 6
Percent Similarity: 46.099 Percent Identity: 24.113
```

alignment\_block:

US-09-540-234-1/rev x AAG47926 ..

Align seg 1/1 to: AAG47926 from: 1 to: 289

```
435 ACACGAGTACATGGCTAGAGTAGTACACTACAAAGG..... 397
:::|||||:::|||||:::|||||:::
163 SerArgAlaTyrGluLeuAspLeuValProAlaGluThrSerIleLysVa 179
396 .....GAAACGAAGTACTGTAGTACACAGGTGTCGCCAGACCAGG 357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
179 lclYasnGlyGluLysLysIleLysSerThrTyrLeuSer..... 192
356 CAGCAAAACAGACTGCTGCCAGGACGATACAGCAAGGAAGTCTCGAGTG 307
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
193 .....CysLeuaspAsnLysLysGlnGlnMetGluArgVal 204
306 CTGCACGGTAAACACACTCAAAATGGCTCCAAAGCGGAGTACAGCAAGAA 257
:::|||||:::
205 PheTyrGlyGln.....ArgG1 210
256 GTCATCAGCGCTACACAGTAACAGCACAGCAATGGCTCCGATTAACATC 207
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
210 uGlnArgMetAlaThr.....HisGluValValArgSerHisA 223
206 GAGCGCATTTGGGTTGTAGTCTGTCTACAACTACGGCTTCACCCAGCTC 157
|||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
223 rGArgGlu.....ProGluValArgAspAspGlyTrpMetGluIle 236
156 TCGAGGGCGATGTTTGAGACG..... 136
::: ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 GluLeuGlyGluPheGluThrGlySerGlyGluGlyAspAspLysG1 253
135 .ACCCTCAAGGGCAGGACTGCTTTTCATATATCCGTCGCCAGCTACAG 87
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
253 uValValMetSerLeuThrGluValLysGlyTyrGlnLeuLysGlyGlyI 270
86 TAGCCGTCGCGCTTGGAGTG 64
:::|||||:::|||||:::|||||:::|||||:::
270 leAlaIleaspGlyIleGluVal 277
```

seq\_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT: AAY85574

seq\_documentation\_block:

ID: AAY85574 standard; Protein; 2608 AA.

```
XX AAY85574;
XX AC
XX DT
XX 07-JUL-2000 (first entry)
XX Hs-UNC-53/3/GFP fusion insert of plasmid pGI3303.
XX UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
KW anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
```

```
KW antisclerotic; antimetastatic; anti-arthritic; autoimmune disease.
XX Synthetic.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Misc-difference 1194 /note= "unspecified"
XX WO9963080-A1.
XX 09-DEC-1999.
XX 02-JUN-1999; 99WO-EP03848.
XX 03-JUN-1998; 98GB-0011962.
XX (JANC ) JANSSEN PHARM NV.
XX Luyten WHML, De Raeymaeker MC, Geysen JJGH, Bogaert TAOE;
XX Maerten LJS, Verhasselt P, Van De Craen M;
XX WPI; 2000-116370/10.
XX N-PSDB; AAA07846.
XX Novel proteins and nucleic acids e.g. for treating neurodegeneration -
XX Disclosure; Fig 7e; 146pp; English.
XX The invention provides vertebrate (human) protein homologue of a UNC-53
XX protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
XX their plus ends. The UNC-53 sequences are used to promote neural
XX regeneration, revascularization and wound healing; also for treating
XX neurodegenerative disease, acute traumatic injury, fibrotic disease and
XX autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
XX polynucleotides can be used for recombinant production of the proteins,
XX as a source of probes for detecting allelic variants and polymorphisms,
XX for sequencing genomic DNA and for detecting UNC-53 expression; and as
XX source of therapeutic antisense sequences. Cells that express the
XX protein are used to identify regulators of cell shape, growth, motility
XX and migration. They can also be used to identify proteins that are
XX involved in signal transduction pathways also involving UNC-53, and to
XX identify compounds that alter attachment of UNC-53 to microtubules. A
XX target gene coupled to a UNC-53 encoding sequence may be used to deliver
XX the target gene to a cellular microtubule or its plus ends. The present
XX sequence represents the amino acid sequence of a full-length Hs-UNC-53/3
XX in fusion with GFP insert of plasmid pGI3305.
```

XX Sequence 2608 AA;

```
alignment_scores:
  Quality: 75.00      Length: 140
  Ratio: 1.027       Gaps: 7
Percent Similarity: 52.143 Percent Identity: 30.714
```

alignment\_block:

US-09-540-234-1/rev x AAY85574 ..

Align seg 1/1 to: AAY85574 from: 1 to: 2608

```
415 AGTAGTACACTACAAAGGGAACGAAGTACTG.....TAGTACACAGG 372
|||||:::|||||::: ||||| |||
755 SerLysThrThrAlaAlaLysLysGluSerLeuIleProSerSerSerG1 771
371 TGTCCCCAGACGACGACG.....AACAGACTATGCTGCG 337
|:::|||||:::|||||::: ||||| |||
771 yileProLysProGlySerLysValProThrValLysGlnThrIleSerP 788
336 CAGCGAGTACAGCAAGGAAGTCTCGAGTGTGCGACGCTAAACACACTC 287
|||||:::|||||::: ||||| |||
788 roGlySerThrAlaSerLysGluSerGluLysPheArgThrThr..... 802
```

```
286 AAATGGCTCCAGCGGAGTACAGCAAGCAAGTCAATCAAGCGCTACACAGT 237
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
803 .....LysGlySerGlnSerLeuSerLysProIleThrMe 816
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
236 AAACAGCACACGATGCTCGATACATACGACGCCAATGGCGTTGTT.A 188
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
816 tGlulys.....AlaSerAlaSerSerCysProAlaProLeuGluG 830
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
187 GTCTTGTAACAACTACGCTCCAGCGAGTCTCGAGGGGCGATGTTTGAGA 138
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
830 lyArgGluAlaGlyGlnAlaSerProSer.....GlySerCysThrMet 844
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
137 CGA.....CCCTCAAAGGCGAGGCGACTGTTTTCAT 106
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
845 ThrValAlaGlnSerSerGlyGlnSerThrGlyAsnGlyAlaValGln.. 860
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
105 TATGCCGTCGACGCTACAGTACCGCTGTCGGCTGGAGGTGCTTTGTC 56
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
861 .LeuProGlnGlnGlnGlnHisSerHisProAsnThrAlaThrValAlaP 877
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
55 CTTTCCTTCAGATCACCCAC 36
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
877 rOPheIleTyArgAlaHis 883
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::

seq_name: /StDS8/gcgdata/geneseq/geneseq/AA2000.DAT.AAG06844
seq_documentation_block:
ID AAG06844 standard; Protein; 155 AA.
XX
AC AAG06844;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 3768.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
```



XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.

```
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157855.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 73.00      Length: 128
  Ratio: 0.936       Gaps: 5
  Percent Similarity: 60.938   Percent Identity: 24.219

alignment_block:
US-09-540-234-1 x AAG06843  ..

Align seg 1/1 to: AAG06843 from: 1 to: 225

2  GGTCCAGCGGCCCTTATGTATGCAGGCGCCTAGTGGTCTATCTGAAGG 51
   |||  :::::  :::::  :::::  :::::  :::::  :::::
94  GlySerTyrValLeuLeuPheLeuAlaGlyLysGlyGlyHisThrHis 110
   :  :::::  :::::  :::::  :::::  :::::  :::::
52  AAAGGACAAAGACACCTCCAGCGGACAGCGGTACTGTAGCTGGCAGCG 101
   :  :::::  :::::  :::::  :::::  :::::  :::::
110 rHisnGlnProMetGluLys.....MetAlaValAlaGlyLeuV 124
   :  :::::  :::::  :::::  :::::  :::::  :::::
102 CATAATGAAACACATCCCTGCCCTTTTGAGGGTCTCTCAACATGCC 151
   :  :::::  :::::  :::::  :::::  :::::  :::::
124 aL.....LeuValProAlaLeuSerProCysAlaThrThrLeuPro 137
   :  :::::  :::::  :::::  :::::  :::::  :::::
152 CTCGAGACGCTCGGTGACGCCGTAGTTTGTACAGACTAACACGCCAATG 201
   :  :::::  :::::  :::::  :::::  :::::  :::::
138 Val.....PheLeuAlaValGlyAsnSerLys 147
   :  :::::  :::::  :::::  :::::  :::::  :::::
202 GCGTCGATGTTATCGGACCATTCGTGTGCTTTTACTGTGTAGCGCTTG 251
   :  :::::  :::::  :::::  :::::  :::::  :::::
147 erMetMetValLeuAlaIleValLeuLeuPheSerThrIleLeuVal 163
   :  :::::  :::::  :::::  :::::  :::::  :::::
252 ATGACTTCTTCTGTACTCCGCTTGGAGCCATTGAGTGTGTTTACCG 301
   |||||  :::::  :::::  :::::  :::::  :::::
164 MetThrSerLeuValAlaLeuSerPheTyrGlyAlaSerGlnLeuLysPh 180
   :  :::::  :::::  :::::  :::::  :::::  :::::
302 TGCAGCAGCTCGAAGACTTCTTGTGCTACTCGCTGCAGCATAGTCTGTT 351
   :  :::::  :::::  :::::  :::::  :::::  :::::
180 eHisTrpValGluArgTyrAspLysLeuLeuValGlySerValLeuCysL 197
   :  :::::  :::::  :::::  :::::  :::::  :::::
```

```
352 TGCTGCCTGGTCTGGGGACACCTGTGTACTAC 383
   |||:  |||::  |||  :::::  :::::
197 euVal...GlyIleLeuThrLeuLeuPheHis 206

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT: AAY01519

seq_documentation_block:
ID_ AAY01519 standard; Protein; 1462 AA.
XX  AAY01519;
AC
XX  14-JUN-1999 (first entry)
DT
XX  A carcinogenesis-inhibiting protein.
DE
XX  Carcinogenesis-inhibiting activity; genetic treatment; cerebral tumour.
KW
XX  Homo sapiens.
OS
XX  JP11075844-A.
PN
XX  23-MAR-1999.
PD
XX  01-SEP-1997; 97JP-0236208.
PF
XX  01-SEP-1997; 97JP-0236208.
PR
XX  (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA
XX  WPI; 1999-257694/22.
DR
XX  N-PSDB; X026546.
DR
XX  A carcinogenesis-inhibiting gene - useful for genetic treatment of
PT  cerebral tumours
XX
XX  Claim 1; Page 15-19; 21pp; Japanese.
PS
XX  The present sequence represents a protein with carcinogenesis-inhibiting
CC  activity. The gene is useful for the genetic treatment of cerebral
CC  tumours.
CC
XX  Sequence 1462 AA;
SQ

alignment_scores:
  Quality: 72.50      Length: 96
  Ratio: 1.318       Gaps: 5
  Percent Similarity: 57.292   Percent Identity: 28.125

alignment_block:
US-09-540-234-1/rev x AAY01519  ..

Align seg 1/1 to: AAY01519 from: 1 to: 1462

443 TCGAACAGACAGCAGGTACATG.....GCTAAGAAGTAGTAC...CA 406
   |||||  :::::  :::::  :::::  :::::  :::::
801 SerAsnValArgGluTyrGlnLysAlaArgAlaLysLysLysTyIleGl 817
   :  :::::  :::::  :::::  :::::  :::::  :::::
405 CTACAAAGGGAACGAAGTACTGTAGTACAGGTGTCCTCCAGACGAGC 356
   :  :::::  :::::  :::::  :::::  :::::  :::::
817 uHisArgSerAsnGluThrSerValIleHisSerLeuProPheGlyGlu 834
   :  :::::  :::::  :::::  :::::  :::::  :::::
355 AGCAACAGACACTATGCTCCAGCGAGTACAGCAAGGAAGTCTTCGAGTGC 306
   |||||  :::::  :::::  :::::  :::::  :::::
834 InThrPheArgProArgGlyMetLeuTyr.....GluCysGlnGluCys 848
   :  :::::  :::::  :::::  :::::  :::::  :::::
305 .....TGCACGGTAAACAGCACTCAATGGCTCCAAGCGGAGTACAGCA 262
   |||||  :::::  :::::  :::::  :::::  :::::
849 GlyGluCysPheAlaHisSerSerAspLeuThr..... 859
   :  :::::  :::::  :::::  :::::  :::::  :::::
261 AGGAGTATCATCAGCGCTACACAGTAAACAGCACACGATGCTCGGATA 212
   :  :::::  :::::  :::::  :::::  :::::  :::::
```



PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134258.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 10-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.



```
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 09-OCT-1999; 99US-0158369.
PR 10-OCT-1999; 99US-0158493.
PR 11-OCT-1999; 99US-0158637.
PR 12-OCT-1999; 99US-0158794.
PR 13-OCT-1999; 99US-0158929.
PR 14-OCT-1999; 99US-0159130.
PR 15-OCT-1999; 99US-0159295.
PR 16-OCT-1999; 99US-0159431.
PR 17-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0159741.
PR 19-OCT-1999; 99US-0159898.
PR 20-OCT-1999; 99US-0160045.
PR 21-OCT-1999; 99US-0160191.
PR 22-OCT-1999; 99US-0160338.
PR 23-OCT-1999; 99US-0160485.
PR 24-OCT-1999; 99US-0160632.
PR 25-OCT-1999; 99US-0160779.
PR 26-OCT-1999; 99US-0160926.
PR 27-OCT-1999; 99US-0161073.
PR 28-OCT-1999; 99US-0161220.
PR 29-OCT-1999; 99US-0161367.
PR 30-OCT-1999; 99US-0161514.
PR 31-OCT-1999; 99US-0161661.
PR 01-NOV-1999; 99US-0161808.
PR 02-NOV-1999; 99US-0161955.
PR 03-NOV-1999; 99US-0162102.
PR 04-NOV-1999; 99US-0162249.
PR 05-NOV-1999; 99US-0162396.
PR 06-NOV-1999; 99US-0162543.
PR 07-NOV-1999; 99US-0162690.
PR 08-NOV-1999; 99US-0162837.
PR 09-NOV-1999; 99US-0162984.
PR 10-NOV-1999; 99US-0163131.
PR 11-NOV-1999; 99US-0163278.
PR 12-NOV-1999; 99US-0163425.
PR 13-NOV-1999; 99US-0163572.
PR 14-NOV-1999; 99US-0163719.
PR 15-NOV-1999; 99US-0163866.
PR 16-NOV-1999; 99US-0164013.
PR 17-NOV-1999; 99US-0164160.
PR 18-NOV-1999; 99US-0164307.
PR 19-NOV-1999; 99US-0164454.
PR 20-NOV-1999; 99US-0164601.
PR 21-NOV-1999; 99US-0164748.
PR 22-NOV-1999; 99US-0164895.
PR 23-NOV-1999; 99US-0165042.
PR 24-NOV-1999; 99US-0165189.
PR 25-NOV-1999; 99US-0165336.
PR 26-NOV-1999; 99US-0165483.
PR 27-NOV-1999; 99US-0165630.
PR 28-NOV-1999; 99US-0165777.
PR 29-NOV-1999; 99US-0165924.
PR 30-NOV-1999; 99US-0166071.
PR 01-DEC-1999; 99US-0166218.
PR 02-DEC-1999; 99US-0166365.
PR 03-DEC-1999; 99US-0166512.
PR 04-DEC-1999; 99US-0166659.
PR 05-DEC-1999; 99US-0166806.
PR 06-DEC-1999; 99US-0166953.
PR 07-DEC-1999; 99US-0167100.
PR 08-DEC-1999; 99US-0167247.
PR 09-DEC-1999; 99US-0167394.
PR 10-DEC-1999; 99US-0167541.
PR 11-DEC-1999; 99US-0167688.
PR 12-DEC-1999; 99US-0167835.
PR 13-DEC-1999; 99US-0167982.
PR 14-DEC-1999; 99US-0168129.
PR 15-DEC-1999; 99US-0168276.
PR 16-DEC-1999; 99US-0168423.
PR 17-DEC-1999; 99US-0168570.
PR 18-DEC-1999; 99US-0168717.
PR 19-DEC-1999; 99US-0168864.
PR 20-DEC-1999; 99US-0169011.
PR 21-DEC-1999; 99US-0169158.
PR 22-DEC-1999; 99US-0169305.
PR 23-DEC-1999; 99US-0169452.
PR 24-DEC-1999; 99US-0169599.
PR 25-DEC-1999; 99US-0169746.
PR 26-DEC-1999; 99US-0169893.
PR 27-DEC-1999; 99US-0170040.
PR 28-DEC-1999; 99US-0170187.
PR 29-DEC-1999; 99US-0170334.
PR 30-DEC-1999; 99US-0170481.
PR 31-DEC-1999; 99US-0170628.
PR 01-JAN-2000; 99US-0170775.
PR 02-JAN-2000; 99US-0170922.
PR 03-JAN-2000; 99US-0171069.
PR 04-JAN-2000; 99US-0171216.
PR 05-JAN-2000; 99US-0171363.
PR 06-JAN-2000; 99US-0171510.
PR 07-JAN-2000; 99US-0171657.
PR 08-JAN-2000; 99US-0171804.
PR 09-JAN-2000; 99US-0171951.
PR 10-JAN-2000; 99US-0172098.
PR 11-JAN-2000; 99US-0172245.
PR 12-JAN-2000; 99US-0172392.
PR 13-JAN-2000; 99US-0172539.
PR 14-JAN-2000; 99US-0172686.
PR 15-JAN-2000; 99US-0172833.
PR 16-JAN-2000; 99US-0172980.
PR 17-JAN-2000; 99US-0173127.
PR 18-JAN-2000; 99US-0173274.
PR 19-JAN-2000; 99US-0173421.
PR 20-JAN-2000; 99US-0173568.
PR 21-JAN-2000; 99US-0173715.
PR 22-JAN-2000; 99US-0173862.
PR 23-JAN-2000; 99US-0174009.
PR 24-JAN-2000; 99US-0174156.
PR 25-JAN-2000; 99US-0174303.
PR 26-JAN-2000; 99US-0174450.
PR 27-JAN-2000; 99US-0174597.
PR 28-JAN-2000; 99US-0174744.
PR 29-JAN-2000; 99US-0174891.
PR 30-JAN-2000; 99US-0175038.
PR 31-JAN-2000; 99US-0175185.
PR 01-FEB-2000; 99US-0175332.
PR 02-FEB-2000; 99US-0175479.
PR 03-FEB-2000; 99US-0175626.
PR 04-FEB-2000; 99US-0175773.
PR 05-FEB-2000; 99US-0175920.
PR 06-FEB-2000; 99US-0176067.
PR 07-FEB-2000; 99US-0176214.
PR 08-FEB-2000; 99US-0176361.
PR 09-FEB-2000; 99US-0176508.
PR 10-FEB-2000; 99US-0176655.
PR 11-FEB-2000; 99US-0176802.
PR 12-FEB-2000; 99US-0176949.
PR 13-FEB-2000; 99US-0177096.
PR 14-FEB-2000; 99US-0177243.
PR 15-FEB-2000; 99US-0177390.
PR 16-FEB-2000; 99US-0177537.
PR 17-FEB-2000; 99US-0177684.
PR 18-FEB-2000; 99US-0177831.
PR 19-FEB-2000; 99US-0177978.
PR 20-FEB-2000; 99US-0178125.
PR 21-FEB-2000; 99US-0178272.
PR 22-FEB-2000; 99US-0178419.
PR 23-FEB-2000; 99US-0178566.
PR 24-FEB-2000; 99US-0178713.
PR 25-FEB-2000; 99US-0178860.
PR 26-FEB-2000; 99US-0179007.
PR 27-FEB-2000; 99US-0179154.
PR 28-FEB-2000; 99US-0179301.
PR 29-FEB-2000; 99US-0179448.
PR 30-FEB-2000; 99US-0179595.
PR 01-MAR-2000; 99US-0179742.
PR 02-MAR-2000; 99US-0179889.
PR 03-MAR-2000; 99US-0179936.
PR 04-MAR-2000; 99US-0180083.
PR 05-MAR-2000; 99US-0180230.
PR 06-MAR-2000; 99US-0180377.
PR 07-MAR-2000; 99US-0180524.
PR 08-MAR-2000; 99US-0180671.
PR 09-MAR-2000; 99US-0180818.
PR 10-MAR-2000; 99US-0180965.
PR 11-MAR-2000; 99US-0181112.
PR 12-MAR-2000; 99US-0181259.
PR 13-MAR-2000; 99US-0181406.
PR 14-MAR-2000; 99US-0181553.
PR 15-MAR-2000; 99US-0181700.
PR 16-MAR-2000; 99US-0181847.
PR 17-MAR-2000; 99US-0181994.
PR 18-MAR-2000; 99US-0182141.
PR 19-MAR-2000; 99US-0182288.
PR 20-MAR-2000; 99US-0182435.
PR 21-MAR-2000; 99US-0182582.
PR 22-MAR-2000; 99US-0182729.
PR 23-MAR-2000; 99US-0182876.
PR 24-MAR-2000; 99US-0183023.
PR 25-MAR-2000; 99US-0183170.
PR 26-MAR-2000; 99US-0183317.
PR 27-MAR-2000; 99US-0183464.
PR 28-MAR-2000; 99US-0183611.
PR 29-MAR-2000; 99US-0183758.
PR 30-MAR-2000; 99US-0183905.
PR 31-MAR-2000; 99US-0184052.
PR 01-APR-2000; 99US-0184199.
PR 02-APR-2000; 99US-0184346.
PR 03-APR-2000; 99US-0184493.
PR 04-APR-2000; 99US-0184640.
PR 05-APR-2000; 99US-0184787.
PR 06-APR-2000; 99US-0184934.
PR 07-APR-2000; 99US-0185081.
PR 08-APR-2000; 99US-0185228.
PR 09-APR-2000; 99US-0185375.
PR 10-APR-2000; 99US-0185522.
PR 11-APR-2000; 99US-0185669.
PR 12-APR-2000; 99US-0185816.
PR 13-APR-2000; 99US-0185963.
PR 14-APR-2000; 99US-0186110.
PR 15-APR-2000; 99US-0186257.
PR 16-APR-2000; 99US-0186404.
PR 17-APR-2000; 99US-0186551.
PR 18-APR-2000; 99US-0186698.
PR 19-APR-2000; 99US-0186845.
PR 20-APR-2000; 99US-0186992.
PR 21-APR-2000; 99US-0187139.
PR 22-APR-2000; 99US-0187286.
PR 23-APR-2000; 99US-0187433.
PR 24-APR-2000; 99US-0187580.
PR 25-APR-2000; 99US-0187727.
PR 26-APR-2000; 99US-0187874.
PR 27-APR-2000; 99US-0188021.
PR 28-APR-2000; 99US-0188168.
PR 29-APR-2000; 99US-0188315.
PR 30-APR-2000; 99US-0188462.
PR 01-MAY-2000; 99US-0188609.
PR 02-MAY-2000; 99US-0188756.
PR 03-MAY-2000; 99US-0188903.
PR 04-MAY-2000; 99US-0189050.
PR 05-MAY-2000; 99US-0189197.
PR 06-MAY-2000; 99US-0189344.
PR 07-MAY-2000; 99US-0189491.
PR 08-MAY-2000; 99US-0189638.
PR 09-MAY-2000; 99US-0189785.
PR 10-MAY-2000; 99US-0189932.
PR 11-MAY-2000; 99US-0190079.
PR 12-MAY-2000; 99US-0190226.
PR 13-MAY-2000; 99US-0190373.
PR 14-MAY-2000; 99US-0190520.
PR 15-MAY-2000; 99US-0190667.
PR 16-MAY-2000; 99US-0190814.
PR 17-MAY-2000; 99US-0190961.
PR 18-MAY-2000; 99US-0191108.
PR 19-MAY-2000; 99US-0191255.
PR 20-MAY-2000; 99US-0191402.
PR 21-MAY-2000; 99US-0191549.
PR 22-MAY-2000; 99US-0191696.
PR 23-MAY-2000; 99US-0191843.
PR 24-MAY-2000; 99US-0191990.
PR 25-MAY-2000; 99US-0192137.
PR 26-MAY-2000; 99US-0192284.
PR 27-MAY-2000; 99US-0192431.
PR 28-MAY-2000; 99US-0192578.
PR 29-MAY-2000; 99US-0192725.
PR 30-MAY-2000; 99US-0192872.
PR 31-MAY-2000; 99US-0193019.
PR 01-JUN-2000; 99US-0193166.
PR 02-JUN-2000; 99US-0193313.
PR 03-JUN-2000; 99US-0193460.
PR 04-JUN-2000; 99US-0193607.
PR 05-JUN-2000; 99US-0193754.
PR 06-JUN-2000; 99US-0193901.
PR 07-JUN-2000; 99US-0194048.
PR 08-JUN-2000; 99US-0194195.
PR 09-JUN-2000; 99US-0194342.
PR 10-JUN-2000; 99US-0194489.
PR 11-JUN-2000; 99US-0194636.
PR 12-JUN-2000; 99US-0194783.
PR 13-JUN-2000; 99US-0194930.
PR 14-JUN-2000; 99US-0195077.
PR 15-JUN-2000; 99US-0195224.
PR 16-JUN-2000; 99US-0195371.
PR 17-JUN-2000; 99US-0195518.
PR 18-JUN-2000; 99US-0195665.
PR 19-JUN-2000; 99US-0195812.
PR 20-JUN-2000; 99US-0195959.
PR 21-JUN-2000; 99US-0196106.
PR 22-JUN-2000; 99US-0196253.
PR 23-JUN-2000; 99US-0196400.
PR 24-JUN-2000; 99US-0196547.
PR 25-JUN-2000; 99US-0196694.
PR 26-JUN-2000; 99US-0196841.
PR 27-JUN-2000; 99US-0196988.
PR 28-JUN-2000; 99US-0197135.
PR 29-JUN-2000; 99US-0197282.
PR 30-JUN-2000; 99US-0197429.
PR 01-JUL-2000; 99US-0197576.
PR 02-JUL-2000; 99US-0197723.
PR 03-JUL-2000; 99US-0197870.
PR 04-JUL-2000; 99US-0198017.
PR 05-JUL-2000; 99US-0198164.
PR 06-JUL-2000; 99US-0198311.
PR 07-JUL-2000; 99US-0198458.
PR 08-JUL-2000; 99US-0198605.
PR 09-JUL-2000; 99US-0198752.
PR 10-JUL-2000; 99US-0198899.
PR 11-JUL-2000; 99US-0199046.
PR 12-JUL-2000; 99US-0199193.
PR 13-JUL-2000; 99US-0199340.
PR 14-JUL-2000; 99US-0199487.
PR 15-JUL-2000; 99US-0199634.
PR 16-JUL-2000; 99US-0199781.
PR 17-JUL-2000; 99US-0199928.
PR 18-JUL-2000; 99US-0200075.
PR 19-JUL-2000; 99US-0200222.
PR 20-JUL-2000; 99US-0200369.
PR 21-JUL-2000; 99US-0200516.
PR 22-JUL-2000; 99US-0200663.
PR 23-JUL-2000; 99US-0200810.
PR 24-JUL-2000; 99US-0200957.
PR 25-JUL-2000; 99US-0201104.
PR 26-JUL-2000; 99US-0201251.
PR 27-JUL-2000; 99US-0201398.
PR 28-JUL-2000; 99US-0201545.
PR 29-JUL-2000; 99US-0201692.
PR 30-JUL-2000; 99US-0201839.
PR 31-JUL-2000; 99US-0201986.
PR 01-AUG-2000; 99US-0202133.
PR 02-AUG-2000; 99US-0202280.
PR 03-AUG-2000; 99US-0202427.
PR 04-AUG-2000; 99US-0202574.
PR 05-AUG-2000; 99US-0202721.
PR 06-AUG-2000; 99US-0202868.
PR 07-AUG-2000; 99US-0203015.
PR 08-AUG-2000; 99US-0203162.
PR 09-AUG-2000; 99US-0203309.
PR 10-AUG-2000; 99US-0203456.
PR 11-AUG-2000; 99US-0203603.
PR 12-AUG-2000; 99US-0203750.
PR 13-AUG-2000; 99US-0203897.
PR 14-AUG-2000; 99US-0204044.
PR 15-AUG-2000; 99US-0204191.
PR 16-AUG-2000; 99US-0204338.
PR 17-AUG-2000; 99US-0204485.
PR 18-AUG-2000; 99US-0204632.
PR 19-AUG-2000; 99US-0204779.
PR 20-AUG-2000; 99US-0204926.
PR 21-AUG-2000; 99US-0205073.
PR 22-AUG-2000; 99US-0205220.
PR 23-AUG-2000; 99US-0205367.
PR 24-AUG-2000; 99US-0205514.
PR 25-AUG-2000; 99US-0205661.
PR 26-AUG-2000; 99US-0205808.
PR 27-AUG-2000; 99US-0205955.
PR 28-AUG-2000; 99US-0206102.
PR 29-AUG-2000; 99US-0206249.
PR 30-AUG-2000; 99US-0206396.
PR 31-AUG-2000; 99US-0206543.
PR 01-SEP-2000; 99US-0206690.
PR 02-SEP-2000; 99US-0206837.
PR 03-SEP-2000; 99US-0206984.
PR 04-SEP-2000; 99US-0207131.
PR 05-SEP-2000; 99US-0207278.
PR 06-SEP-2000; 99US-0207425.
PR 07-SEP-2000; 99US-0207572.
PR 08-SEP-2000; 99US-0207719.
PR 09-SEP-2000; 99US-0207866.
PR 10-SEP-2000; 99US-0208013.
PR 11-SEP-2000; 99US-0208160.
PR 12-SEP-2000; 99US-0208307.
PR 13-SEP-2000; 99US-0208454.
PR 14-SEP-2000; 99US-0208601.
PR 15-SEP-2000; 99US-0208748.
PR 16-SEP-2000; 99US-0208895.
PR 17-SEP-2000; 99US-0209042.
PR 18-SEP-2000; 99US-0209189.
PR 19-SEP-2000; 99US-0209336.
PR 20-SEP-2000; 99US-0209483.
PR 21-SEP-2000; 99US-0209630.
PR 22-SEP-2000; 99US-0209777.
PR 23-SEP-2000; 99US-0209924.
PR 24-SEP-2000; 99US-0210071.
PR 25-SEP-2000; 99US-0210218.
PR 26-SEP-2000; 99US-0210365.
PR 27-SEP-2000; 99US-0210512.
PR 28-SEP-2000; 99US-0210659.
PR 29-SEP-2000; 99US-0210806.
PR 30-SEP-2000; 99US-0210953.
PR 01-OCT-2000; 99US-0211100.
PR 02-OCT-2000; 99US-0211247.
PR 03-OCT-2000; 99US-0211394.
PR 04-OCT-2000; 99US-0211541.
PR 05-OCT-2000; 99US-0211688.
PR 06-OCT-2000; 99US-0211835.
PR 07-OCT-2000; 99US-0211982.
PR 08-OCT-2000; 99US-0212129.
PR 09-OCT-2000; 99US-0212276.
PR 10-OCT-2000; 99US-0212423.
PR 11-OCT-2000; 99US-0212570.
PR 12-OCT-2000; 99US-0212717.
PR 13-OCT-2000; 99US-0212864.
PR 14-OCT-2000; 99US-0213011.
PR 15-OCT-2000; 99US-0213158.
PR 16-OCT-2000; 99US-0213305.
PR 17-OCT-2000; 99US-0213452.
PR 18-OCT-2000; 99US-0213599.
PR 19-OCT-2000; 99US-0213746.
PR 20-OCT-2000; 99US-0213893.
PR 21-OCT-2000; 99US-0214040.
PR 22-OCT-2000; 99US-0214187.
PR 23-OCT-2000; 99US-0214334.
PR 24-OCT-2000; 99US-0214481.
PR 25-OCT-2000; 99US-0214628.
PR 26-OCT-2000; 99US-0214775.
PR 27-OCT-2000; 99US-0214922.
PR 28-OCT-2000; 99US-0215069.
PR 29-OCT-2000; 99US-0215216.
PR 30-OCT-2000; 99US-0215363.
PR 31-OCT-2000; 99US-0215510.
PR 01-NOV-2000; 99US-0215657.
PR 02-NOV-2000; 99US-0215804.
PR 03-NOV-2000; 99US-0215951.
PR 04-NOV-2000; 99US-0216098.
PR 05-NOV-2000; 99US-0216245.
PR 06-NOV-2000; 99US-0216392.
PR 07-NOV-2000; 99US-0216539.
PR 08-NOV-2000; 99US-0216686.
PR 09-NOV-2000; 99US-0216833.
PR 10-NOV-2000; 99US-0216980.
PR 11-NOV-2000; 99US-0217127.
PR 12-NOV-2000; 99US-0217274.
PR 13-NOV-2000; 99US-0217421.
PR 14-NOV-2000; 99US-0217568.
PR 15-NOV-2000; 99US-0217715.
PR 16-NOV-2000; 99US-0217862.
PR 17-NOV-2000; 99US-0218009.
PR 18-NOV-2000; 99US-0218156.
PR 19-NOV-2000; 99US-0218303.
PR 20-NOV-2000; 99US-0218450.
PR 21-NOV-2000; 99US-0218597.
PR 22-NOV-2000; 99US-0218744.
PR 23-NOV-2000; 99US-0218891.
PR 24-NOV-2000; 99US-0219038.
PR 25-NOV-2000; 99US-0219185.
PR 26-NOV-2000; 99US-0219332.
PR 27-NOV-2000; 99US-0219479.
PR 28-NOV-2000; 99US-0219626.
PR 29-NOV-2000; 99US-0219773.
PR 30-NOV-2000; 99US-0219920.
PR 01-DEC-2000; 99US-0220067.
PR 02-DEC-2000; 99US-0220214.
PR 03-DEC-2000; 99US-0220361.
PR 04-DEC-2000; 99US-0220508.
PR 05-DEC-2000; 99US-0220655.
PR 06-DEC-2000; 99US-0220802.
PR 07-DEC-2000; 99US-0220949.
PR 08-DEC-2000; 99US-0221096.
PR 09-DEC-2000; 99US-0221243.
PR 10-DEC-2000; 99US-0221390.
PR 11-DEC-2000; 99US-0221537.
PR 12-DEC-2000; 99US-0221684.
PR 13-DEC-2000; 99US-0221831.
PR 14-DEC-2000; 99US-0221978.
PR 15-DEC-2000; 99US-0222125.
PR 16-DEC-2000; 99US-0222272.
PR 17-DEC-2000; 99US-0222419.
PR 18-DEC-2000; 99US-0222566.
PR 19-DEC-2000; 99US-0222713.
PR 20-DEC-2000; 99US-0222860.
PR 21-DEC-2000; 99US-0223007.
PR 22-DEC-2000; 99US-0223154.
PR 23-DEC-2000; 99US-0223301.
PR 24-DEC-2000; 99US-0223448.
PR 25-DEC-2000; 99US-0223595.
PR 26-DEC-2000; 99US-0223742.
PR 27-DEC-2000; 99US-0223889.
PR 28-DEC-2000; 99US-0224036.
PR 29-DEC-2000; 99US-0224183.
PR 30-DEC-2000; 99US-0224330.
PR 31-DEC-2000; 99US-0224477.
PR 01-JAN-2001; 99US-0224624.
PR 02-JAN-2001; 99US-0224771.
PR 03-JAN-2001; 99US-0224918.
PR 04-JAN-2001; 99US-0225065.
PR 05-JAN-2001; 99US-0225212.
PR 06-JAN-2001; 99US-0225359.
PR 07-JAN-2001; 99US-0225506.
PR 08-JAN-2001; 99US-0225653.
PR 09-JAN-2001; 99US-0225800.
PR 10-JAN-2001; 99US-0225947.
PR 11-JAN-2001; 99US-0226094.
PR 12-JAN-2001; 
```











Align seg 1/1 to: AAR52971 from: 1 to: 3969

```

367 CCAGACGAGCAGCAACAGACTATGCTCCGAGGAGTACAGCAAGGAA 318
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3406 ProGlnLeuGlyThrSerGlnThrProSerThrAlaAlaIleThrAlaA 3422
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 GTCTTCAGTGTGTCAGCGGTAACAGACACTCAA..... 285
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3422 aSerSer...IleCysValLeuProSerThrGlnThrThrGlyIleThrAl 3438
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
284 ..ATGGCTCCCAAGCGGA.....GTAC 266
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3438 aAlaSerProSerGlyGluAlaaspGluHisTyrGlnLeuGlnHisVala 3455
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 AGCAA.....GGAAGTCATCAAGCGCTACACAGTAAACAGCAC..... 228
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3455 snGlnLeuLeuAlaSerLysThrGlyIleHisSerSerGlnArgaspLeu 3471
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
227 .....ACGAATGGCTCCGATACATCGACGC 202
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3472 AspSerAlaSerGlyProGlnValSerAsnPheThrGlnThrValaspAl 3488
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 CAVTGGGCTGTAGTCTGTACAA.....ACTA 173
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3488 aProAsnSerMetClyLeuGluGlnAsnLysAlaLeuSerSerAlaValG 3505
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
172 CGGCTTCACGACGCTCTCGAGGGCGCATGTTTGAGACGACGCCCTCAAAGGG 123
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3505 lnAlaSerProThrSerProGlyClySerProSerSerProSerSerGly 3521
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
122 CAG 120
|||
3522 Gln 3522

```

seq\_name: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT: AAR86250

seq\_documentation\_block:

```

ID AAR86250 standard; Protein; 229 AA.
XX
AC AAR86250;
XX
KW Single chain gonadotropin analogue 4.
XX
DE Single chain gonadotropin analogue 4.
XX
KW Single chain gonadotropin; human chorionic gonadotropin; hCG;
KW alpha; beta; subunit; analogue; glycoprotein hormone; fertility;
KW inhibit; stimulate; increase; lutropin; luteinising hormone; LH;
KW follicle stimulating hormone; FSH; vaccine; contraceptive.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= leader
FT Region 19..129
FT /label= hFSH_beta_subunit_(1-111)
FT Region 130..137
FT /label= linker
FT Region 138..229
FT /label= Gonadotropin_alpha_subunit_(1-92)
XX
PN WO9522340-A1.
XX
PD 24-AUG-1995.
XX
PF 17-FEB-1995; 95WO-US02067.
XX
PR 18-FEB-1994; 94US-0199382.
XX
PA (SENS-) SENSE-TEST.
XX
PI Moyle WR;

```

XX  
DR WPI; 1995-302553/39.  
DR N-PSDB; AAT03224.  
XX  
PT Methods for altering fertility in mammals, esp. humans - e.g.  
PT stimulating fertility by reducing the activity and/or levels of  
PT circulating glyco:protein hormones having lutropin activity  
XX  
XX Example 15; Fig 9; 102pp; English.  
XX  
CC Analogue 4 (human FSH-beta(1-111)-linker-human-alpha(1-92)) is a  
CC specific example of a single chain gonadotropin; chimeric proteins  
CC having a chorionic gonadotropin (CG) beta-subunit at the N-terminus  
CC and a CG alpha-subunit at the C-terminus, joined by a linker of 1-16  
CC amino acids are claimed. The analogue has follicle stimulating hormone  
CC (folliotropin) activity and is useful for inducing follicle development  
CC and increasing male fertility.  
XX  
SQ Sequence 229 AA;

alignment\_scores:  
Quality: 70.00 Length: 112  
Ratio: 1.228 Gaps: 5  
Percent Similarity: 50.893 Percent Identity: 25.000

alignment\_block:

US-09-540-234-1 x AAR86250 ..

Align seg 1/1 to: AAR86250 from: 1 to: 229

```

76 GGACACGGCTACTGTAGCTGGCAGCATATGGAACACAGTCCTGCC 125
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer..... 131
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 TTTTGAGGTCGTCTCAACATGCCCTCGAGAGCGTGGTGAAGCCCTAG 175
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 .....GlySerGlySerGlySerAlaProAspValGlnAspCysProG 146
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
176 TTTGTACAAGACTAACACGCCAATGCGTCGATGTTATCGAGCCATTC 225
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
146 luCystThrLeuGlnGluAsnProPheSerGlnProGlyAlaProIle 162
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 GTGTGCTGTTTACTGTGTAGCGTGTGATGACTTCTTCTGCTGT..... 267
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 LeuGlnCys.....MetClyCysCysPheSerAr 172
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
268 .....ACTCCGCTTGGAGCCATTGAGTGTGTTTACCGTGCAGC 307
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
172 gAlaTyrProThrProLeu.....Arg 180
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
308 ACTCGAGAGACTTCTTCTGTAC...TCGCTGGCAGCATAGTCTGTTGC 354
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 erLysLysThrMetLeuValGlnLysAsnValThrSerGluSerThrCys 196
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 TGCCTGCTCTGGGACACCTGTGTACTACAGTACTT 390
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
197 CysValAlaLysSerTyrAsnArgValThrValMet 208
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq\_name: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT: AAR86251

seq\_documentation\_block:

```

ID AAR86251 standard; Protein; 237 AA.
XX
AC AAR86251;
XX
DT 26-APR-1996 (first entry)
XX
DE Single chain gonadotropin analogue 5.
XX
KW Single chain gonadotropin; human chorionic gonadotropin; hCG;
KW alpha; beta; subunit; analogue; glycoprotein hormone; fertility;
KW inhibit; stimulate; increase; lutropin; luteinising hormone; LH;
KW

```

follicle stimulating hormone; FSH; vaccine; contraceptive.

Synthetic.

Key	Location/Qualifiers
Peptide	1..20
Region	/label= leader
Region	21..113
Misc-difference	/label= hCG_beta_subunit_(1-93)
Region	/note= "Arg corresponds to CCG codon"
Region	114..137
Region	/label= hFSH_beta_subunit_(88-111)
Region	138..145
Region	/label= linker
Region	146..237
Region	/label= Gonadotropin_alpha_subunit_(1-92)







```
PS Claim 2; Page 1021; 1453pp; English.
XX AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA2575941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC the invention can be used as vaccines, antibodies and compositions of
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria meningitidis (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria meningitidis bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX Sequence 135 AA;
SQ

alignment_scores:
    Quality: 69.00      Length: 122
    Ratio: 1.232      Gaps: 8
Percent Similarity: 45.902 Percent Identity: 28.689

alignment_block:
US-09-540-234-1 x AA25296 ..

Align seg 1/1 to: AA25296 from: 1 to: 135

60 AAGACACCTCCAAAGCGGACACGG...CTACTGTAGCTGGCAGCGCATAA 106
25 ArgValSerProSerThrThrArgTrpMetLeu.AlaTrpSerGlyGluI 41
107 TGGAAACAGCTCCCTGCCCTTTTGGGGTGGTCTCAACATGCCCTCGA 156
41 leSerAlaSerPro..... 45
157 GAGCTGGTCAAGCCGTAGTTTGTACAACTAACACGCCAATGGCGTC 206
46 .....SerAlaAlaLeuAlaThrArgValSerLysArgThrArgAr 59
207 GATGTTATCGGACCATTCGTGTGCTGT.....TTACTGTGTA 244
59 gLeuProSerAlaAlaValCysGlyAspAlaGluLeuLeuCys 76
245 GCCTTGATGACTTCCTTGTCTGCTACTCCGCTTGGAGCCATTTCAGTG... 291
76 erAla.....ThrValSerGlyValProMetThrAlaGluMetValSer 90
292 .....CTGTTTACCGTGCAGCACTCGAAGACTTCCTTT 323
91 SerAlaCysArgArgArgLeuPheArgAla.....ThrSerCy 103
324 GCTGTACTCGTCGACGACATAGTCTGTTTGTCTGCTGGTGGGACACC 373
103 sMetSerSerSerAlaAla.....CysMetSerPheTrpGlyMetI 117
374 TGTGTACTACAGTA 387
117 leCysAlaSerVal 121

seq_name: /SID88/gcgdata/geneseq/genesep/AA2000.DAT:AA25297
seq_documentation_block:
ID AA25297 standard; Protein; 135 AA.
XX
XX AA25297;
XX
XX 21-MAR-2000 (first entry)
XX
XX Neisseria meningitidis ORF 643 protein sequence SEQ ID NO:2068.
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW
```

```
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
XX WO9957280-A2.
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX WPI: 2000-062150/05.
XX N-PSDB; AA254059.
XX
XX Novel Neisseria meningitidis predicted to be useful antigens for
XX vaccines and diagnostics -
XX Claim 2; Page 1022; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA2575941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisseria meningitidis (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria meningitidis bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 135 AA;
SQ

alignment_scores:
    Quality: 69.00      Length: 122
    Ratio: 1.232      Gaps: 8
Percent Similarity: 45.902 Percent Identity: 28.689

alignment_block:
US-09-540-234-1 x AA25297 ..

Align seg 1/1 to: AA25297 from: 1 to: 135

60 AAGACACCTCCAAAGCGGACACGG...CTACTGTAGCTGGCAGCGCATAA 106
25 ArgValSerProSerThrThrArgTrpMetLeu.AlaTrpSerGlyGluI 41
107 TGGAAACAGCTCCCTGCCCTTTTGGGGTGGTCTCAACATGCCCTCGA 156
41 leSerAlaSerPro..... 45
157 GAGCTGGTCAAGCCGTAGTTTGTACAACTAACACGCCAATGGCGTC 206
46 .....SerAlaAlaLeuAlaThrArgValSerLysArgThrArgAr 59
```





FT Misc-difference 50 /note= "wild-type Asn at position 30 of the beta-subunit is pref. replaced by another amino acid (esp. Gln) to remove a glycosylation site"

FT Misc-difference 70

FT Misc-difference 98 /note= "Arg corresponds to CCG codon"

FT Misc-difference 99 /note= "wild-type Pro at position 78 of the beta-subunit is replaced by another amino acid to agree with the glycosylation site motif"

FT Misc-difference 99 /note= "wild-type Val at position 79 of the beta-subunit is replaced by Thr to agree with the glycosylation site motif"

FT Region 114..137

FT /label= hFSH\_beta\_subunit\_(88-111)

FT Region 138..145

FT /label= linker

FT Region 146..237

FT /label= Gonadotropin\_alpha\_subunit\_(1-92)

FT Misc-difference 197 /note= "wild-type Asn at position 52 of the alpha-subunit is pref. replaced by another amino acid (esp. Gln) to remove a glycosylation site"

FT Misc-difference 223 /note= "wild-type Asn at position 78 of the alpha-subunit is pref. replaced by another amino acid (esp. Gln) to remove a glycosylation site"

FT WO9522340-A1.

XX 24-AUG-1995.

XX 17-FEB-1995; 95WO-US02067.

XX 18-FEB-1994; 94US-0199382.

XX (SENS-) SENS1-TEST.

XX Moyle WR;

XX WFI; 1995-302553/39.

XX Methods for altering fertility in mammals, esp. humans - e.g. stimulating fertility by reducing the activity and/or levels of circulating glyco:protein hormones having lutropin activity

XX Example 25; Fig 10 and Page 60; 102pp; English.

XX The single-chain gonadotropin analogue 5b (human CG-beta(1-93) [N13X,N30X,P78X,V79T]-hFSH-beta(88-111)-linker-human CG-alpha(1-92) [N52X,N78X]) is an example of a chimeric glycopeptide hormone having an extra glycosylation site. Addition of oligosaccharides has a positive effect on stability of hormones in circulation and can be used to prevent unwanted antibody or receptor interactions. The present analogue has anti-follicle stimulating hormone (follictropin) activity and is useful for treating ovarian hyperstimulation and reducing spermatogenesis.

XX Sequence 237 AA;

alignment\_scores:  
Quality: 69.00 Length: 112  
Ratio: 1.211 Gaps: 5  
Percent Similarity: 50.893 Percent Identity: 25.000

alignment\_block:  
US-09-540-234-1 x AAR86273 ..

Align seg 1/1 to: AAR86273 from: 1 to: 237

76 GGACACGGCTACTGTAGCTGGCAGCGCATATGGAACACAGTCCCTGCC 125  
||| ::::::::::::::: ::| ::::::::::|  
126 GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer..... 139  
126 TTTTGGGGTCTCTCAACATGCCCTCGAGACGTCGGTGAAGCCGCTAG 175  
||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|  
140 .....GlySerGlySerGlySerAlaProAspValGlnAspCysProG 154  
176 TTTGTACAGACTACACAGCCCAATGGCGTCGATGTTATCGAGCCATTC 225  
||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|  
154 luCysThrLeuGlnGluAsnProPheSerGlnProGlyAlaProile 170  
226 GTGTGCTGTTTACTGTAGCGCTGTGACTCTCTTCTGCTG..... 267  
||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|  
171 LeuGlnCys.....MetGlyCysCysPheSerAr 180  
268 .....ACTCGCTTGGAGCCATTGTAGTGTCTTTACCGTGCAGC 307  
||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|  
180 gAlaTyrProThrProLeu.....ArgS 188  
308 ACTCGAGACTTCTGTCTGTAC...TCGCTGGCAGCATAGTCTGTTTGC 354  
||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|  
188 erLysLysThrMetLeuValGlnLys\*\*\*ValThrSerGluSerThrCys 204  
355 TGCCTGGTCTGGGACACCTGTGTACTACACTT 390  
||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|  
205 CysValAlaLysSerTyrAsnArgValThrValMet 216  
seq\_name: /SIDS8/gcgdata/geneseq/AA1995.DAT:AAR86274

seq\_documentation\_block:  
ID AAR86274 standard; Protein; 237 AA.  
XX AAR86274;  
XX 13-MAY-1996 (first entry)  
XX Single chain gonadotropin analogue 6b with extra glycosylation site.  
XX Single chain gonadotropin; human chorionic gonadotropin; hCG;  
KW alpha; beta; subunit; analogue; glycoprotein hormone; fertility;  
KW inhibit; stimulate; increase; lutropin; luteinising hormone; LH;  
KW follicle stimulating hormone; FSH; vaccine; contraceptive.  
XX Synthetic.  
OS OS  
FH Key Location/Qualifiers  
FH Peptide 1..20  
FH /label= leader  
FH Region 21..120  
FH /label= hCG\_beta\_subunit\_(1-100)  
FH Misc-difference 33 /note= "wild-type Asn at position 13 of the beta-subunit is pref. replaced by another amino acid (esp. Gln) to remove a glycosylation site"  
FH Misc-difference 50 /note= "wild-type Asn at position 30 of the beta-subunit is pref. replaced by another amino acid (esp. Gln) to remove a glycosylation site"  
FH Misc-difference 70 /note= "Arg corresponds to CCG codon"  
FH Misc-difference 98 /note= "wild-type Pro at position 78 of the beta-subunit is replaced by another amino acid to agree with the glycosylation site motif"  
FH Misc-difference 99 /note= "wild-type Val at position 79 of the beta-subunit is replaced by Thr to agree with the glycosylation site motif"











236	AAA.....CAGCAC.....	228
	:::	
68	LYSGLYCYSGLUHISLSCYSALACYSARGASNGLYGLYLEUCYSH1A1 84	
227	ACGAATGCTCCGATAAC.....	210
	:::	
84	afthrAsnGLYSerCysSerCysProLeuGLYtrpMetGLYProHisCysG 101	
209	.....ATCGACGCCATTGGCGTTGTAGTCTTGTCACAAACT 174	
	:::     :::	
101	luHisAlaCysProAlaGLYArgTYrGLYAlaAlaCysLeuLeuGLuCys 117	
173	.....ACGGTTCACCGAGCTCTCGAGGGCAGTGT 142	
	:::     :::     :::	
118	SerCysGlnAsnAsnGLYSerCysGluProthrSer...GlyAlaCysLe 133	
141	GAGACGACCCCTCA.....	129
	:::	
133	uCysGLYProGLYpHeTYrGLYGLnAlaCysGLuAspThrCysProAlaG 150	
128	....AAAGGCGAGGAGCTGTTTCCATTATFCGCTGCCAGCTACAGTAg 84	
	:::     :::	
150	LYPheHisGLYSerGLYCysGlnArgValCysGLuCysGlnGLnGLYAla 166	
83	CGGTGTCGC.....GCTTGAGGAGTGCTTTGTCTT 54	
	:::     :::     :::	
167	ProCysAspProValSerGLYArgCysLeuCysPro 178	

CC cancer and neurological diseases. The proteins may be used to stimulate  
CC the growth and motility of keratinocytes, to inhibit the growth of  
CC cancer cells, to modulate angiogenesis and tumour vascularisation, to  
CC modulate skin inflammation, to modulate epithelial cell growth and to  
CC inhibit binding of HIV-1 to leukocytes. The invention may also be used  
CC to treat growth and developmental defects, skin wounds and hair follicle  
CC disorders. Sequences AAY75942-V76123 represent polypeptides encoded  
CC by cDNA sequences derived from several mouse, rat or human skin cell  
CC types. Sequences AAY75942-V75947, AAY76020-V76021, AAY76094-V76104 and  
CC AAY76119 are proteins with an N-terminal signal sequence, indicating  
CC that they are secreted. Sequences AAY75986-V75989, AAY76061-V76071,  
CC AAY76106-V76109 and AAY76121-V76122 are proteins with one or more  
CC putative transmembrane domains.

seq_documentation_block:	
XX ID	AAAY76014 standard; Protein; 299 AA.
XX AC	AAAY76014;
XX XX	27-MAR-2000 (first entry)
DT XX	
XX DE	Murine EGF family protein, SEQ ID NO:192.
XX XX	Skin; dermal papilla; keratinocyte; neonatal foreskin fibroblast;
KW KW	embryonic skin cell; keratinocyte stem cell; transit amplifying cell;
KW KW	screted; transmembrane; inflammation; cancer; neurological disease;
KW KW	angiogenesis; tumour vascularisation; growth disorder;
KW KW	developmental disorder; skin wound; hair follicle disorder;
KW KW	anti-inflammatory; cystostatic; neuroprotective; vulvetry.
XX OS	Mus sp.
XX OS	WO9955865-A1.
PN XX	
XX XX	04-NOV-1999.
XX PD	
XX XX	29-APR-1999; 99WO-NZ00051.
XX PF	
XX PR	29-APR-1998; 98US-0069726.
PR PR	09-NOV-1998; 98US-0188930.
XX XX	(GENE-) GENESIS RES & DEV CORP LTD.
XX PA	
PI PI	Strachan L, Sleeman M, Watson JD, Onrust R, Kumble A, Murison JG;
XX XX	
DR DR	WPI: 2000-072177/06.
DR N	PSDB; AAZ61678.
XX XX	
PT PT	Novel polynucleotides useful for the treatment of various conditions
PT PT	including wounds and cancer -
XX XX	
PS PS	Claim 4; Page 132; 235pp; English.
XX XX	
CC CC	The invention relates to novel nucleic acid sequences derived from rat
CC CC	dermal papilla, human keratinocytes and neonatal foreskin fibroblasts,
CC CC	and mouse embryonic skin, keratinocyte stem cells and transit amplifying
CC CC	cells. Polypeptides of the invention may be used to treat inflammation.





**THIS PAGE BLANK (USPTO)**

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_organelle:Q9MR87	+	90.00	161.89	0.2463	353	Q9mr87 melosira ambigua, cytoch
sp_organelle:Q9TE67	+	87.00	156.00	0.5248	353	Q9te67 rhizosolenia setigera, c
sp_bacteria:O83713	-	85.50	156.18	0.7662	236	O83713 treponema pallidum, cons
sp_invertebrate:O161717	+	85.50	147.84	0.7655	689	O161717 caenorhabditis elegans
sp_invertebrate:Q9V714	-	84.50	140.94	0.9845	1298	Q9v714 drosophila melanogast
sp_invertebrate:Q9NGV2	-	84.50	140.57	0.9844	1361	Q9ngv2 drosophila melanogast
sp_organelle:Q9TE70	+	84.00	150.10	1.12	353	Q9te70 ditylum brightwellii, cy
sp_organelle:Q97466	+	82.00	146.17	1.85	353	Q97466 eustigmatos majus, cyto
sp_organelle:Q97471	+	82.00	146.17	1.85	353	Q97471 ophiocytus magnus, cyto
sp_organelle:O21352	+	79.00	142.22	3.94	275	O21352 euhadra herklotzi, cyto
sp_organelle:Q99384	+	79.00	140.27	3.94	353	Q99384 compomenia bullosa, cyto
sp_organelle:Q9TE69	+	78.50	139.29	4.47	353	Q9te69 fragilaria striatula, cy
sp_invertebrate:Q9NC83	+	78.00	144.04	5.08	169	Q9nc83 strongylocentrotus pur
sp_organelle:Q99381	+	78.00	138.31	5.07	353	Q99381 pseudochorda nagaii, cy
sp_organelle:Q99382	+	78.00	138.31	5.07	353	Q99382 undaria pinnatifida, cy
sp_organelle:Q99387	+	78.00	137.22	5.07	406	Q9m987 terebratulina pacifica,
sp_organelle:Q9TP94	+	78.00	135.39	5.07	513	Q9tp94 terebratulina retusa, cy
sp_organelle:O21002	+	78.00	135.36	5.07	515	O21002 branchiostoma lanceolat
sp_plant:Q96397	+	78.00	133.67	5.07	640	Q96397 chlamydomonas reinhardt
sp_invertebrate:Q94936	+	77.50	133.10	5.75	607	Q94936 caenorhabditis elegans
sp_virus:Q90368	-	77.50	122.03	5.75	2513	Q90368 o'nyong-nyong virus, n
sp_organelle:Q99385	+	77.00	136.34	6.53	353	Q99385 tribonema marinum, cyto
sp_organelle:Q97467	+	77.00	136.34	6.53	353	Q97467 heterococcus caespitosus
sp_plant:Q9LJ36	+	77.00	134.04	6.53	474	Q9lj36 oryza sativa (rice), hyp
sp_organelle:P92800	+	77.00	133.20	6.53	528	P92800 pylaella littoralis, cy
sp_organelle:Q9TE66	+	76.50	135.36	7.41	353	Q9te66 skeletonema costatum, cy
sp_organelle:Q9TE64	+	76.50	135.36	7.41	353	Q9te64 thalassiosira nordenski
sp_organelle:Q9G407	+	76.50	132.48	7.40	511	Q9g407 lithobius forficatus, cy
sp_organelle:Q9MQ04	+	76.00	136.61	8.40	265	Q9mq04 tetraelmis aff. maculat
sp_invertebrate:O16955	-	76.00	134.46	8.40	349	O16955 caenorhabditis elegans
sp_human:Q9P234	-	76.00	129.76	8.40	638	Q9p234 homo sapiens (human), k
sp_human:Q9HAM2	-	76.00	128.91	8.40	712	Q9ham2 homo sapiens (human), c
sp_organelle:P92619	+	75.50	136.56	9.53	235	P92619 chorella vulgaris, cyto
sp_organelle:Q9TE71	+	75.50	133.39	9.53	353	Q9te71 cyllindrotheca closterium
sp_plant:Q23578	+	75.50	132.76	9.53	383	Q23578 arabidopsis thaliana (mc
sp_organelle:Q63273	+	75.00	133.08	10.81	324	Q63273 cataglyphis velox, cyto
sp_organelle:Q9M393	+	75.00	131.32	10.81	406	Q9m393 japonica nipponica, cyto
sp_organelle:Q9XP16	+	75.00	131.24	10.81	410	Q9xp16 macroplitis maturus, cyt
sp_organelle:Q9G4C3	+	75.00	129.65	10.81	503	Q9g4c3 thraustochytrium aureum,
sp_invertebrate:O17047	-	74.50	126.02	12.25	707	O17047 caenorhabditis elegans

```
123 exSerHisSerGlyProAlaVal.AspLeuAlaIlePheSerLeuHisLe 139
194 CGCAATGCGTCGATGTTATCGGAGCCATTGCTGTGCTGTTTACTGTGT 243
139 uSerGlyAlaIleSerLeuGlyAlaIleAsnPheIleCysThrIleL 156
244 AGCGTGTGATGACTTCCTGCTGCTACTCC.....GCT 275
156 euAsnMetArgThrLysGlyLeuPheMetHisLysLeuProLeuValVal 172
276 TGGAGCCATTGAGTCTGCTTTACCGTCGACGACCTCGAAGACTTCCTG 325
173 TrpSerIleLeuIleThrAlaValLeuLeuLeuLeuSer.....LeuPr 187
326 TGTACTCGCTGCACATAGTCTGTTGCTG..... 356
187 oValLeuAlaGlyAlaIleThrMetLeuLeuLeuThrAspArgAsnPheAsnT 204
357 .....CCTGCTCGGACACACCTGCTGTACTACAGTACT 389
204 hrThrPhePheAspProAlaGlyGlyAspProValLeuTyrGlnHis 220
390 TCGTTT.....CCCTTTGTAGTGTACTTCTT..... 419
221 LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProAl 237
420 .....AGCCATGTACTCGTGTCT 437
237 apheGlyIleIleSerHisIleIleValSer 247
```

seq\_name: sp\_organelle:Q9TE67

```
seq_documentation_block:
ID Q9TE67 PRELIMINARY; PRT; 353 AA.
AC Q9TE67;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COXI.
OS Rhizosolenia setigera.
OG Mitochondrion.
OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC Rhizosoleniophycidae; Rhizosoleniales; Rhizosoleniaceae; Rhizosolenia.
OX NCBI_TaxID=3005;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCMP 1330;
RA Ehara M., Inagaki Y., Watanabe K.I., Ohama T.;
RT "Phylogenetic analysis of diatom cox1 genes and implications of
RT fluctuating GC content on mitochondrial genetic code evolution.";
RL Curr. Genet. 0:0-0(1999).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERRICYTOCHROME C (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AB020226; BAA86611.1; -.
DR InterPro; IPR000883; -.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PR01165; CYCOXIDASE1.
DR PROSITE; PS00077; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
```

```
FT NON_TER 1 1
FT NON_TER 353 353
SQ SEQUENCE 353 AA; 38138 MW; 65D9C11AA4D48870 CRC64;

alignment_scores:
  Quality: 87.00      Length: 161
  Ratio: 1.160      Gaps: 8
  Percent Similarity: 46.584      Percent Identity: 27.329

alignment_block:
US-09-540-234-1 x Q9TE67 ..
Align seg 1/1 to: Q9TE67 from: 1 to: 353

107 TGGAAACAGTCCTCCCTTTTGGGGTCGTCACAAACATGCCCTCGA 156
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
90 TrpLeuLeuProProSerLeuLeuValAlaSerMetLeuAlaG 106
157 GAGC.....TCGGTGAAGCCGTAGTTTGTACAGACT 188
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
106 uAlaGlyValGlyThrGlyTrpThrValTyrPro.ProLeuSerSerGly 122
189 AACAAAGCCAAATGGCTCGATGTT..... 212
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
123 AsnSerHisSerGlyAlaSerValAspLeuAlaIlePheSerLeuHisLe 139
213 .....ATCGGAGCCATTGCTGTGCTGCTGTTTACTGTGT 243
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
139 uSerGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleCysThrIleL 156
244 AGCGTGTGATGACTTCCTGCTGCTACTCC.....GCT 275
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
156 euAsnMetArgValLysGlyLeuPheMetHisArgLeuProLeuPheVal 172
276 TGGAGCCATTGAGTCTGCTTTACCGTCGACGACCTCGAAGACTTCCTGC 325
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
173 TrpSerIleLeuIleThrAlaValLeuLeuLeuLeuSer.....LeuPr 187
326 TGTACTCGCTGCACATAGTCTGTTGCTG..... 356
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
187 oValLeuAlaGlyAlaIleThrMetLeuLeuLeuThrAspArgAsnPheAsnT 204
357 .....CCTGCTCGGACACACCTGCTGTACTACAGTACT 389
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 hrThrPhePheAspProAlaGlyGlyAspProValLeuTyrGlnHis 220
390 TCGTTT.....CCCTTTGTAGTGTACTTCTT..... 419
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
221 LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProG 237
420 .....AGCCATGTACTCGTGTCT 437
237 yPheGlyIleIleSerHisIleIleValSer 247

seq_name: sp_bacteria:O83713
seq_documentation_block:
ID O83713 PRELIMINARY; PRT; 236 AA.
AC O83713;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TP0731.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
```



237 .....TACCGTGCAGCATCGAGAGACTTCTTCCTG 324  
195 rAlaThrTyrTysGlnHisArgAlaAlaValTyrSerLeuIleAla 212  
328 TACTCGCTGCGCAGCATAGCTGTGTTGCTGCTGGGACACCTGTG 377  
212 lnPheAlaThrSerIleIleCysPheSerPro.....ProIle 224  
378 TACTACAGTACTTCGGTTCCTTGTAGTGGTACTACTTCTTAGCCATGT 427  
225 phe.....LeuValPheValAlaPhePheGlnLeuProHisal 237



alignment\_scores:  
 Quality: 84.50 Length: 109  
 Ratio: 1.482 Gaps: 5  
 Percent Similarity: 52.294 Percent Identity: 28.440

alignment\_block:  
 US-09-540-234-1/rev x Q9NGV2 ..

Align seg 1/1 to: Q9NGV2 from: 1 to: 1361

```

365 CAGACAGGCGAGCAACAGACTACTGTCGCCAGCGAGTACAGCAAGGAAGT 316
||||| |||||||: : : : : : : : : : : : : : : : : : : : : : :
302 GlnThrTyrGlnGlnThrSerAsnThrAsnThrAlaLysAsnArgArgLy 318
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
315 CTTTCAGTGTGTCAGCGTAACAGCACTCAATGGCTCCAGCGGAGTAC 266
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
318 sTyrProAsnValThrSerAsnLysValGlnMet..... 329
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
265 AGCAAGAACTCATCAAGCGCTACACAGTAAACAGACAGCAATGGCTCC 216
||||| : : : : : : : : : : : : : : : : : : : : : : :
330 .....HisIleThrArgGluThrAsnArgLeu 338
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
215 GATAACATCAGCCGCTT.....GGCGTTGTTAGTCTTGTACAACTAC 172
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
339 GluAsnIleSerAlaAlaGluProAlaValValGlyAlaAlaMetThrTh 355
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
171 G.....GCTTCACCGACGCTCTCAGGGGCGATGTTTGAGACGACCCCTCA 128
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
355 rAspArgSerThrProThrCysAsnLeuAspCys.....G 367
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
127 AAGGGCAGGAGCTGTTTTCATTATGCGCGTCCAGCTACAGTACGCCGTGT 78
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
367 lySerAspGlyIle.....CysAlaLeuGluAlaThrAla 378
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
77 CCGGCTGGAGGTGCTTTGCTCTTC 51
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
379 AlaSerSerArgCysLeuCysProPhe 387
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

seq\_name: sp\_organelle:Q9TE70

seq\_documentation\_block:  
 ID Q9TE70 PRELIMINARY; PRT; 353 AA.  
 AC Q9TE70;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2001 (TrEMBLrel. 16, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).  
 GN COXI.  
 OS Ditylum brightwellii.  
 OG Mitochondrion.  
 OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;  
 OC Lithodemiaceae; Ditylum.  
 OX NCBI\_TaxID=49249;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CCAP1022-2;  
 RA Ehara M., Inagaki Y., Watanabe K.I., Ohama T.;  
 RT "Phylogenetic analysis of diatom coxi genes and implications of  
 RT fluctuating GC content on mitochondrial genetic code evolution.";  
 RL Curr. Genet. 0:0-0(1999).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
 CC FERRICYTOCHROME C (BY SIMILARITY).  
 CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 DR EMBL; AB020223; BAA86608.1; -.  
 DR InterPro; IPR000883; -.  
 DR Pfam; PF00115; COX1; 1.  
 DR PRINTS; PR01165; CYCOXIDASEI.  
 DR PROSITE; PS00077; COX1; 1.  
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
 KW Respiratory chain; Transmembrane.  
 FT NON\_TER 1 353  
 FT SEQUENCE 353 AA; 38145 MW; CC5018EC930C9685 CRC64;  
 SQ

alignment\_scores:  
 Quality: 84.00 Length: 145  
 Ratio: 1.105 Gaps: 8  
 Percent Similarity: 52.414 Percent Identity: 28.276

alignment\_block:  
 US-09-540-234-1 x Q9TE70 ..

Align seg 1/1 to: Q9TE70 from: 1 to: 353

```

92 GCTGGCAGCGCAATAATGAAACAGTCCCTGCCCCCTTTGAGGGTCGCTC 141
|||||: : : : : : : : : : : : : : : : : : : : : : :
109 AlaGlyThrGly...TrpThrValTyrPro.ProLeu.....SerSerA 122
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
142 AATACGCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAGACTAAC 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 laThrAlaHisSerGlyGlyAlaValAspLeuAlaIlePheSerLeuHis 138
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
192 AAGCCCAATGGCGTCGATGTTATCGAGCCATTCGTCGTGTTACTGT 241
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
139 LeuSerGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleCysThrI 155
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
242 GTAGCGCTTCATGACT.....TCCTTGCTGCTACTCCG 273
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
155 ePheAsnMetArgValLysSerLeuSerPheHisAsnLeuProLeuPheV 172
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
274 CTTGGAGCCATTTCAGTGTGTTTACCGTCGACGACCTCGAAGACTTCCT 323
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 alTrpSerValLeuIleThrAlaPheLeuLeuLeuSer.....Leu 186
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
324 GCTGTACTCGCTGGCAGCATAGTCTGTTGCTG..... 356
|||||: : : : : : : : : : : : : : : : : : : : : : :
187 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 203
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
357 .....CCTGGTCTGGGACACCTGTGTGTACTACAGTA 387
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
203 nThrThrPhePheAspProAlaGlyGlyAspProValLeuTyrGlnH 220
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
388 CTTCGTTT.....CCCTTTGTAGTGTACTACTTCTT... 419
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
220 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuLeuPro 236
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 .....AGCCATGTACTCGTGTCT 437
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
237 GlyPheGlyIleValSerHisIleValSer 247
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

seq\_name: sp\_organelle:047466

seq\_documentation\_block:  
 ID 047466 PRELIMINARY; PRT; 353 AA.  
 AC 047466;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).  
 GN COXI.  
 OS Eustigmatos magnus.  
 OG Mitochondrion.  
 OC Eukaryota; stramenopiles; Eustigmatophyceae; Eustigmatos.  
 OX NCBI\_TaxID=73014;

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CCAP 860/2;
RX  MEDLINE=97383241; PubMed=9236270;
RA  Ehara M., Hayashi-Ishimaru Y., Inagaki Y., Ohama T.;
RT  "Use of a deviant mitochondrial genetic code in yellow-green algae as
RL  a landmark for segregating members within the phylum.";
RJ  J. Mol. Evol. 45:119-124(1997).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CCAP 860/2;
RX  MEDLINE=97451042; PubMed=9302324;
RA  Inagaki Y., Hayashi-Ishimaru Y., Ehara M., Inagaki Y., Ohama T.;
RT  "Algae or protozoa: phylogenetic position of euglenophytes and
RL  dinoflagellates as inferred from mitochondrial sequences.";
RJ  J. Mol. Evol. 45:295-300(1997).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CCAP 860/2;
RX  MEDLINE=9800465; PubMed=9342410;
RA  Hayashi-Ishimaru Y., Ehara M., Inagaki Y., Ohama T.;
RT  "A deviant mitochondrial genetic code in pyrenesophytes (yellow-
RL  algae): UGA codon for tryptophan.";
RJ  Curr. Genet. 32:296-299(1997).
CC  -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC  CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT'S 1-
CC  3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC  CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGNATING IN
CC  CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC  AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC  AND COPPER B (BY SIMILARITY).
CC  -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4
CC  FERROCYTOCHROME C.
CC  -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC  -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC  -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC  -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR  EMBL; AB000205; BAA24969.1; -.
DR  InterPro; IPR000883; -.
DR  Pfam; PF00115; COX1; 1.
DR  PROSITE; PS00077; COX1; 1.
KW  Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW  Respiratory chain; Transmembrane.
KW  NON_TER 1
FT  NON_TER 353
SQ  SEQUENCE 353 AA; 37913 MW; PF491AFB46532F00 CRC64;

alignment_scores:
    Quality: 82.00      Length: 146
    Ratio: 1.079       Gaps: 8
    Percent Similarity: 52.055      Percent Identity: 28.767

alignment_block:
US-09-540-234-1 x 047466  ..

Align seg 1/1 to: 047466 from: 1 to: 353

92  GCTGGCAGCATATGGAACAGTCCTGCTGCTTTGAGGGTCTCTC 141
109  AlaGlyThrGly...TrpThrValTyProProLeuSerSerValGlnAl 124
142  AAACATGCCCTCGAGACGTCGGTGAAGCGGTAGTTGTACAAAGACTAAC 191
124  aHisSer.....GlyProSerVal.AspLeuAlaIlePheSerLeuHis 138
192  AACGCCAATGGCGTCGATGTTATCGAGGCCATTCGTGTCTGTTACTGTT 241
139  LeuAlaGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleThrVa 155
242  GTAGCGCTTG.....ATGACTTCCTTCGTGTACTCGG 273
155  lPheAsnMetArgGlyProGlyIleThrMethHisArgLeuProLeuPheV 172

```

```

274  CTTGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
172  aItrpAlaValLeuIleThrAlaPheLeuLeuValLeuSer.....Leu 186
324  GCTGTACTCGTCGGCAGCATAGTCTGTTTCTGCTG..... 356
187  ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 203
357  .....CTGTGCTGGGACACCTGTGTACTACAGTA 387
203  nThrThrPhePheAspProAlaGlyGlyAspProValLeuTyrglnH 220
388  CTTGCTTT.....CCTTTGTAGTGTACTACTTCTT... 419
220  lSeuPheTrpPhePheGlyHisProGluValTyIleLeuIleLeuPro 236
420  .....AGCATGACTGCTGTCTGTT 440
237  AlaPheGlyIleIleSerHisValValSerSerLeu 248

seq_name: sp_organelle:047471

seq_documentation_block:
ID 047471 PRELIMINARY; PRT; 353 AA.
AC 047471;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COXI.
OS Ophiocytium majus.
OC Mitochondrion.
OC Eukaryota; stramenopiles; Xanthophyceae; Ophiocytium.
OX NCBI_TaxID=73019;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCAP 855/1;
RX MEDLINE=97383241; PubMed=9236270;
RA Ehara M., Hayashi-Ishimaru Y., Inagaki Y., Ohama T.;
RT "Use of a deviant mitochondrial genetic code in yellow-green algae as
RL a landmark for segregating members within the phylum.";
RJ J. Mol. Evol. 45:119-124(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CCAP 855/1;
RX MEDLINE=97451042; PubMed=9302324;
RA Inagaki Y., Hayashi-Ishimaru Y., Ehara M., Igarashi I., Ohama T.;
RT "Algae or protozoa: phylogenetic position of euglenophytes and
RL dinoflagellates as inferred from mitochondrial sequences.";
RJ J. Mol. Evol. 45:295-300(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CCAP 855/1;
RX MEDLINE=9800465; PubMed=9342410;
RA Hayashi-Ishimaru Y., Ehara M., Inagaki Y., Ohama T.;
RT "A deviant mitochondrial genetic code in pyrenesophytes (yellow-
RL algae): UGA codon for tryptophan.";
RJ Curr. Genet. 32:296-299(1997).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT'S 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGNATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4
CC FERROCYTOCHROME C.
CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AB000210; BAA24974.1; -.

```





Ratio: 1.172 Gaps: 5  
Percent Similarity: 51.538 Percent Identity: 26.923

## alignment\_block:

US-09-540-234-1 x Q9TE69 ..

Align seg 1/1 to: Q9TE69 from: 1 to: 353

```
135 TCCTGCTCAAAATGCGCCCTCGAGCGCTGCGTAGTTGTGACAA 184
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 SerSerIleThrAlaHisSerGlyGlyAlaValAspLeuAlaIlePheSe 136
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
185 GACTAACAACGCCAATGGCTCGATGCTATTACGGACCATTCGTGCTGT 234
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
136 rLeuHisValSerGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleC 153
235 TTACTGTGTAGCGCTGTGATGACATTCCTTGCTG..... 266
||||| ||||| ||||| ||||| ||||| ||||| |||||
153 yThrIlePheAsnMetArgValIlySerLeuSerPheHisLysLeuPro 169
267 TACTCGCGCTTGGAGCCATTTGAGTGTGTTTACCGTGCAGCAGCTCGAAGA 316
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
170 LeuPheValIrrSerValIleIleThrAlaPheLeuLeuLeuSer.. 185
317 CTTCCTGTGCTACTCGCTGGCAGCATAGCTCTTTGCTG..... 356
||| ||||| ||||| ||||| ||||| ||||| |||||
186 ....LeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgA 201
357 .....CTCGTCTGGGACACCTGTGTAC 380
201 snPheAsnThrThrPhePheAspProAlaGlyGlyGlyAspProValLeu 217
381 TACAGTACTTCGTTT.....CCCTTTGTAGTGTACTA.. 413
||||| ||||| ||||| ||||| ||||| ||||| |||||
218 TyrGlnHisLeuPheThrPhePheGlyHisProGluValTyrIleLeuII 234
414 .....CTTCTTACCCATGCTACTGCTGTCT 437
234 eValProGlyPheGlyIleIleSerHisValIleValSer 247
```

seq\_name: sp\_invertebrate:Q9NC83

## seq\_documentation\_block:

```
ID Q9NC83 PRELIMINARY; PRT; 169 AA.
AC Q9NC83;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 18.3 KDA PROTEIN.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20253941; PubMed=10793471;
RA Rast J.P., Pancer Z., Davidson E.H.;
RT "New approaches towards an understanding of deuterostome immunity.";
RL Curr. Top. Microbiol. Immunol. 248:3-16(2000).
DR EMBL; AF228878; AAF78206.1; -.
KW Hypothetical protein.
SQ SEQUENCE 169 AA; 18272 MW; D5098D9ABBE2B628 CRC64;
```

## alignment\_scores:

Quality: 78.00 Length: 125  
Ratio: 1.322 Gaps: 5  
Percent Similarity: 47.200 Percent Identity: 24.800

## alignment\_block:

US-09-540-234-1 x Q9NC83 ..

Align seg 1/1 to: Q9NC83 from: 1 to: 169

```
64 CACCTCAAGCCGCGACACGCTACTGTAGCTGGCAGCGCATATGGAAGA 113
||||| ||||| ||||| ||||| ||||| ||||| |||||
63 HisIleGlnAla..... 66
114 CAGTCCTCCCTGCCCTTTTGGAGGTGCTCTCAAAACATGCCCTTCGAGACGTCG 163
::: ||||| ||||| ||||| ||||| ||||| |||||
67 GlnProGlnProGlnThrThrValValArgThrAlaProArgSerThrA 83
::: ||||| ||||| ||||| ||||| ||||| |||||
164 GTAAGCCGTAGTTTGTACAAGACTAACACGCAATGGCGTGCATGTTA 213
::: ||||| ||||| ||||| ||||| ||||| |||||
83 snAlaPheSerGlyAlaProArgGlnGlnValProAspTyrSerLeuMet 99
214 TCGAGGCAATTCGTGCTGCTTACTGTGTAGCGCTTGATGACTTCCTCTG 263
::: ||||| ||||| ||||| ||||| ||||| |||||
100 AlaTrpMetValThrLeuCys.....Cy 107
264 CTGTACTCCGCTTGGAGCCATTTGAGTGTGTTTACGTCGACGACATCGA 313
||||| ||||| ||||| ||||| ||||| ||||| |||||
107 sCysLeuProPheGlyLeuVal...AlaValPheMetAlaSerArgAlaL 123
314 AGACTTCCTTCGCTGTAC.....TCGCTGGCAGCATAGTCTGTGTT 351
|| ||||| ||||| ||||| ||||| ||||| |||||
123 ysAspLysGlnMetTyrGlyAspAspGlyAlaArgSerThrSerGln 139
352 TCGTGTGCTGCTGG.....GGACACCTGTGTACTACAGTACTCG 392
||||| ||||| ||||| ||||| ||||| ||||| |||||
140 CysAlaLeuGlyTyrAlaValAlaGlyIleValcysGlyIleValLeuSe 156
393 TTTCCCTTTTGTAGTGTACTACTTC 417
::: ||||| ||||| ||||| ||||| ||||| |||||
156 rIleLeuLeuGlyIleTyrPhe 164
```

seq\_name: sp\_organelle:O99381

## seq\_documentation\_block:

```
ID O99381 PRELIMINARY; PRT; 353 AA.
AC O99381;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COXI.
OS Pseudochorda nagaii.
OC Mitochondrion.
OC Eukaryota; stramenopiles; Phaeophyceae; Laminariales;
OC Pseudochordaceae; Pseudochorda.
OX NCBI_TaxID=74379;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEUMURO, HOKKAIDO, JAPAN;
RA Ehara M., Inagaki Y., Watanabe K.I., Hayashi-Ishimaru Y., Kawai H.,
RA Ohama T.;
RT "Phylogenetic analyses of heterokont algae with emphasis on the AUA
codon reassignment in mitochondria.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC 1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
AND COPPER B (BY SIMILARITY).
CC 1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
FERRICYTOCHROME C.
CC 1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC 1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC 1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AF037992; AAC94978.1; -.
DR InterPro; IPR000883; -.
DR InterPro; IPR002106; -.
```

GN COAT.  
OS Undaria pinnatifida.  
OC Mitochondrion.  
CG Eukaryota; stramenopiles; Phaeophyceae; Laminariales; Alariaceae;  
OC Undaria.  
OX NCBI\_TaxID=74381;  
RN [1]





174 AGTTGTACAGACATACCAACGCGCAATGGCGGAGTGTATCGAGACCA 223

145 A|a|l|l|e|p|h|e|s|e|r|L|e|u|H|s|t|e|u|A|G|I|V|V|S|e|r|S|e|r|I|l|e|G|I|V|A|I|I 162

146 A l a t l e p h e s e r f e u H i s l e u A l a g l v v a l s e r s e r f i l e u c l y a l a i l 162

```

224 TCCTGTCTCTTACTCTGTAGCGCTGATGACTTCTCTCTCTACTCC.. 272
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
162 eAsnPheIleThrThrIleHisAsnMetArgAlaSerIleGluTrpAsnA 179
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
273 .....GCTTGGAGCCATTGAGTCTGCTGTTTACCTGCGACCA 308
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
179 rgValProLeuPheValTrpSerIleTrpValThrAlaTrpLeuLeu 195
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
309 CTCGAAGACTTCCTGTCTGTCTGCTGCGCAGCATAGCTCTGTTTGTG 356
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
196 LeuSer.....LeuProValLeuAlaGlyAlaIleThrMetLeuLeu 210
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
357 .....CCTGGTCTGGGACAC 372
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
210 rAspArgAsnIleAsnThrThrPhePheAspProSerGlyGlyAsp 227
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
373 CTGTGTACTACAGTACTCTGTTT.....CCCTTTGTAGTG 407
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
227 rolleLeuTrpGluHisLeuPheTrpPhePheGlyHisProGluValTr 243
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
408 GTACTACTCTCT.....AGCATGTACTCTGTG 434
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
244 IleLeuLeuProGlyPheGlyIleIleSerHisIleIle 258
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::

```

seq\_name: sp\_plant:Q96397

```

seq_documentation_block:
ID   Q96397  PRELIMINARY;          PRT;   640 AA.
AC   Q96397; O04833; O04733;
DT   01-FEB-1997 (TREMBLrel. 02, Created)
DT   01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE   01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE   LRG5.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gloeckner G., Beck C.F.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Gloeckner G., Beck C.F.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR   EMBL; U73817; AAB17561.1; -.
DR   EMBL; U73818; AAB39840.1; -.
DR   InterPro; IPR001007; -.
DR   PROSITE; PS01208; VWFC; UNKNOWN 1.
SQ   SEQUENCE 640 AA; 67298 MW; 764EEETADC32FC99 CRC64;

```

```

alignment_scores:
  Quality: 78.00      Length: 192
  Ratio: 1.099       Gaps: 8
  Percent Similarity: 36.979  Percent Identity: 21.875

```

alignment\_block:

```

US-09-540-234-1 x Q96397 ..
Align seg 1/1 to: Q96397 from: 1 to: 640

11 GCCTTATGTATGCAGGAGCGCTAGTGGTCATCTGAAGGAAGACAA 60
   |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
427 AlaAlaValProAlaAlaAlaCysSerGlyPheArgGlyGlyG1 443
   |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
61 AGACACTCAACCGGACACGCGCTACTGTAGCTGCACGGCATATGGA 110
   |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
443 yGlyValAlaArg.....ValAlaAlaGlyThrSerArgAlaA 456
   |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
111 AAACAGTCCCTGCCCTTTTGGGGTCTCTCAAAACATGCCCTCGAGACG 160

```

```

|||||  |||
456 laglyValProArgArgLeu..... 462
161 TCGGTGAAGCCGTAGTTTGTACAAAGACTAAACGCAATGGCGTCGA.. 208
    |||||  ::  |||||  ::  |||||  ::  |||||  ::
463 .....GluArgArgTrpArgArgAr 469
208 ..... 208
469 gGlyArgGlyTrpArgArgArgValArgArgArgGlyAlaGlyArgA 486
209 .....TGTTATCGAGCCATTCGTGCTGCTTTACTGTCTAGCGCTTGAT 253
    |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||  ::
486 laValCysThrAlaGlyArgCysCysTrpMetThrCysLeuProMetTrp 502
254 GACTTCCTCTGCTGTACTCCGCTTGGAGCCATTGAG..... 289
    |||||  ::  |||||  ::  |||||  ::  |||||  ::
503 GlySerGlyGlyThrTrpProTrpArgProLeuMetThrProSerArgTh 519
290 ....TGCTGTTTACCGTGCAGCAGCTCGAAGACTTCCTCTGCTGTACT... 331
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::
519 rCysAlaCysLeuPro.....ThrProCysCysSerArgT 531
332 .....CCCTGGCAGCAT 343
    |||||  ::
531 rpLeuArgArgTrpArgCysGlyTrpAlaProGlyGlyArgTrpArgCys 547
344 AGTCTGTTTGTGCTGG.....TCTGGGACACACTGTGTACT 381
    |||||  ::  |||||  ::  |||||  ::  |||||  ::
548 SerLeuCysSerCysTrpArgTrpGlyCysSerGlyArgThrProLeuLe 564
382 ACAGTACTTCGTT.....TCCCTTTGTAGTGGTACTA 413
    ||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::
564 uProThrTrpValTrpArgArgCysCysArgCysCysArgGlySerA 581
414 CTTCCTAGCCATGTACTCTGTCTGT 439
    ::  |||  ::  |||||  ::
581 rgAlaProArgCysAsnTrpValCys 589
seq_name: sp_invertebrate:O45936
seq_documentation_block:
ID   O45936  PRELIMINARY;          PRT;   607 AA.
AC   O45936;
DT   01-JUN-1998 (TREMBLrel. 06, Created)
DT   01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT   01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE   Y43F4B.7 PROTEIN.
DE   Y43F4B.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Matthews L.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE-94150718; PubMed-7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA   Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
    elegans."
RL   Nature 368:32-38(1994).

```

```
DR EMBL; AL021481; CAA16336.1; -.
DR InterPro; IPR002422; -.
DR Pfam; PF01490; Aa_trans; 1.
SQ SEQUENCE 607 AA; 68571 MW; 8C643382F7AED090 CRC64;

alignment_scores:
  Quality: 77.50      Length: 136
  Ratio: 1.062       Gaps: 8
Percent Similarity: 53.676 Percent Identity: 24.265

alignment_block:
US-09-540-234-1 x 045936 ..
Align seg 1/1 to: 045936 from: 1 to: 607

58 CAAGACACTCCCAAGCGGACACGGCTACTGTAGCTGGCAGGCATTAAT 107
   ::::::::::::::: |||::: |||::: |||::: |||:::
266 GluAsnArgMetGlnSerProHisAlaPheIleSerTrpAsnGlyVal 282
108 GGAACACAGTCCCTGC.....:::.....CCTT 127
282 u...AsnSerSerCysLeuValValLeuAlaIlePheSerValThrGlyP 298
128 TTGAGGTCGTCTCAACATGCCCTCGAGACGTGGTGAAGCCGTAGTT 177
298 heTyrGlyTyrLeuSer...LeuGlyAsnAspValLysAspThrAlaThr 313
178 TGTACAAGACTAACACGCCCAATGCGTCGATGTTATCGGACCATTCGT 227
314 LeuAsnLeuProMetThrProPheTyrGlnThrIleLysLeuMetPheVa 330
228 GNGCTGTTTACTGTAGCG.....CTTGATGACTTCTTGTCTGCTACTC 271
330 lAlaCysIleMet.IleSerTyrProLeuGlnPheTyrValProMetGlu 346
272 CGCTGGAGCCATTGAGTGTGTTTACCCTGCAGCAGCTCGAAGACTTCC 321
347 ArgIleGluLysTip.....:::.....IleThrArgLysIlePr 357
322 TTGCTGTACTCGCTGGCAGCATAGTCTGTTTGTCTGCTGCTGGTGGGAC 370
357 oValAspLysGlnThrLeuTyrIleTyrIleAlaArgTyrSerGlyValI 374
371 .....ACCTGTGTA.....:::.....CTACAGTACTTCTGTT 394
374 leLeuThrCysAlaIleAlaGluLeuIleProHisLeuAlaLeuPheIle 390
395 TCCTTT 400
391 SerLeu 392

seq_name: sp_virus:090368

seq_documentation_block:
ID 090368 PRELIMINARY; PRT; 2513 AA.
AC 090368;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE.
OS O'nyong-nyong virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11027;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SG650;
RA Lanciotti R.S., Ludwig M.L., Rwanuna E.B., Lutwama J.J., Kram T.M.,
RA Karabatsos N., Cropp B.C., Miller B.R.;
RT "Emergence of Epidemic O'nyong-nyong Fever in Uganda After a 35 Year
RT Absence:Genetic Characterization of the Virus.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL; AF079456; AAC97204.1; -.
DR InterPro; IPR000606; -.
DR InterPro; IPR001091; -.
DR InterPro; IPR001788; -.
DR InterPro; IPR002589; -.
DR InterPro; IPR002620; -.
DR Pfam; PF00978; RNA_dep_RNapol2; 2.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01661; DUF27; 1.
DR Pfam; PF01707; Peptidase_C9; 1.
DR PROSITE; PS00093; N4_MTASE; UNKNOWN_1.
KW POLYPROTEIN.
FT CHAIN 1 535 NON-STRUCTURAL PROTEIN 1.
FT CHAIN 536 1333 NON-STRUCTURAL PROTEIN 2.
FT CHAIN 1334 1896 NON-STRUCTURAL PROTEIN 3.
FT CHAIN 1897 2513 NON-STRUCTURAL PROTEIN 4.
SQ SEQUENCE 2513 AA; 280219 MW; 1B546AD981E5F2A7 CRC64;

alignment_scores:
  Quality: 77.50      Length: 163
  Ratio: 0.912       Gaps: 11
Percent Similarity: 52.147 Percent Identity: 29.448

alignment_block:
US-09-540-234-1/rev x 090368 ..
Align seg 1/1 to: 090368 from: 1 to: 2513

417 GAAGTAGTACCCTACAAAGGAAACGAAGTACTCT..... 382
   |||::: |||::: |||::: |||::: |||:::
1686 GluAlaValProValProSerSerLeuGluAlaCysAspAlaThrMetAs 1702
381 .....AGTACACAGGTGCCCGACGACGACGACGACGACGACGACGAC 339
1702 pTrpProSerIleAspIleValPro.ThrArgGlnArgSerAspSerPhe 1718
338 GCCACGCGAGTAC...AGCAAGGAAGTCTCGAGTCTGCACCGTAAAC.. 294
   |||::: |||::: |||::: |||::: |||:::
1719 AspSerGluTyrSerSerArgSerAsnIleGlnLeuValThrAlaAspVa 1735
293 .....AGCACTCAATGGTCCCAAGCGGAGTACAGCAAGTGC.. 260
   |||::: |||::: |||::: |||::: |||:::
1735 lHisAlaProMetTyrAlaAsnSerLeuAlaSerSerGly.....G 1749
259 GAAGTCATCAACGCTACACAGTAAACAGCACACCAATGCG..... 219
   |||::: |||::: |||::: |||::: |||:::
1749 lYserValLeuSerSerGluGlnAlaGlnAsnGlyIleMetIle 1765
218 .....TCCGATAACATCGACCCATTGGCGTTGTAGTCTTGTACA 178
   |||::: |||::: |||::: |||::: |||:::
1766 LeuProAspSerGluAspThrAspSerIle.....SerArgValse 1779
177 AACTACGGCTTCACCGAGCTCTCGAGG.....GTGTCGGCTTGGAGGT 149
   |||::: |||::: |||::: |||::: |||:::
1779 rThrProIleAlaProProArgArgArgLeuGlyArgThrIleAsnValT 1796
148 CATGTTTGTAGAGACGCCCTCAAAAGG.....GCAGGGACTGTTTC 109
   |||::: |||::: |||::: |||::: |||:::
1796 hrcysaspGluArgGluGlyLysIleLeuProMetAlaSerAspArgLeu 1812
108 CATTATGCGCGTCCAGCTACAGTACGCC.....GTGTCGGCTTGGAGGT 65
   |||::: |||::: |||::: |||::: |||:::
1813 PheThrAlaLysProTyrThrValAlaLeuGlyValSerThrAlaAspIl 1829
64 GTCCTTGTCTTCTTCATGATGACCCACTAGCGCT 28
   ::: ::: |||::: |||::: |||::: |||:::
1829 eThr...AlaTyrProIleGlnAlaProLeuGlySer 1840

seq_name: sp_organelle:099385
seq_documentation_block:
ID 099385 PRELIMINARY; PRT; 353 AA.
```

```

AC 099385;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COXI.
OS Tribonema marinum.
OC Mitochondrion.
OC Eukaryota; stramenopiles; Xanthophyceae; Tribonematales;
OC Tribonemataceae; Tribonema.
OX NCBI_TaxID=74375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIES 548;
RA Ehara M., Inagaki Y., Watanabe K.I., Hayashi-Ishimaru Y., Kawai H.,
RA Ohama T.;
RT "Phylogenetic analyses of heterokont algae with emphasis on the AUA
RT codon reassignment in mitochondria.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERROCYTOCHROME C.
CC -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AF037996; AAC94982.1; -.
DR InterPro; IPR000883; -.
DR TrEMBL; P0002106; -.
DR Pfam; PF00115; COXI; 1.
DR PRINTS; PR01165; CYCOXIDASE1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS00077; COXI; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT NON_TER 1
FT NON_TER 353
SQ SEQUENCE 353 AA; 38121 MW; 7DDCED749BA4681 CRC64;

```

## alignment\_scores:

```

Quality: 77.00 Length: 143
Ratio: 1.055 Gaps: 8
Percent Similarity: 51.049 Percent Identity: 27.273

```

## alignment\_block:

```

US-09-540-234-1 x 099385 ..
Align seg 1/1 to: 099385 from: 1 to: 353
92 GCTGGCAGCGCAATAAGCAACAGTCGCCCTTTTGGGGTGGCTC 141
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 AlaglyThrGly...TrpThrValTrpProLeuSerGlyIleGlnAl 124
142 AAACATGCCCTCGAGACGTCGGTGAAGCGGTAGTTGTACAACTAAC 191
|||||:|||||:|||||:|||||:|||||:|||||:
124 aHisSer.....GlyProSerVal.AspLeuAlaIlePheSerLeuHis 138
192 AACGCCAATGGCTGATGTATCGGAGCCATCGTGTCTGTACTGT 241
|||||:|||||:|||||:|||||:|||||:|||||:
139 LeuSerGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleThrThr 155
242 GTAGCGGTG.....ATGACTTCCTCTGCTGCTACTCCG 273
|||||:|||||:|||||:|||||:|||||:|||||:
155 ePheAsnMetArgAlaProGlyMetThrMetAspArgLeuProLeuPheV 172
274 CTTGGAGCCATTTGAGTGTCTTTACCGTGCAGCACTCGAAGACTTCCTT 323

```

```

:::||||| ||| :|||:|||||: |||:|||||: |||
172 alTrpSerValLeuIleThrAlaPheLeuLeuLeuLeuSer.....Leu 186
324 GCTGTACTCGCTGCGACCATAGTCTGTTTCTGCTG..... 356
|||||:|||||:|||||:|||||:|||||:|||||:
187 ProValLeuAlaGlyGlyIleThrMetLeuThrAspArgAsnPheAs 203
357 .....CCTGGTCTGGGGACACCTGTGTACTACAGTA 387
|||||:|||||:|||||:|||||:|||||:
203 nThrThrPhePheAspProAlaGlyGlyGlyAspProValLeuTyrGlnH 220
388 CTTGCGTTT.....CCCTTTGTAGTGTACTACTTCTT... 419
|||||:|||||:|||||:|||||:|||||:
220 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuLeuLeuPro 236
420 .....AGCCATGTACTC 431
237 GlyPheGlyIleValSerHisIleLeu 245
seq_name: sp_organelle:O47467
seq_documentation_block:
ID O47467 PRELIMINARY; PRT; 353 AA.
AC O47467;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COXI.
OS Heterococcus caespitosus.
OC Mitochondrion.
OC Eukaryota; stramenopiles; Xanthophyceae; Heterococcus.
OX NCBI_TaxID=55585;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCAP 835/2A;
RX MEDLINE=97383241; PubMed=92362770;
RA Ehara M., Hayashi-Ishimaru Y., Inagaki Y., Ohama T.;
RT "Use of a deviant mitochondrial genetic code in yellow-green algae as
RT a landmark for segregating members within the phylum.";
RL J. Mol. Evol. 45:119-124(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CCAP 835/2A;
RX MEDLINE=97451042; PubMed=9302324;
RA Inagaki Y., Hayashi-Ishimaru Y., Ehara M., Igarashi I., Ohama T.;
RT "Algae or protozoa: phylogenetic position of euglenophytes and
RT dinoflagellates as inferred from mitochondrial sequences.";
RL J. Mol. Evol. 45:295-300(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CCAP 835/2A;
RX MEDLINE=98004465; PubMed=9342410;
RA Hayashi-Ishimaru Y., Ehara M., Inagaki Y., Ohama T.;
RT "A deviant mitochondrial genetic code in prymnesiophytes (yellow-
RT algae): UGA codon for tryptophan.";
RL Curr. Genet. 32:296-299(1997).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERROCYTOCHROME C.
CC -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AB000206; BAA24970.2; -.
DR InterPro; IPR000883; -.
DR TrEMBL; IPR000894; -.

```

DR Pfam; PF00115; COX1; 1.  
 DR PRINTS; PRO1165; CYCOKIDASE1.  
 DR PRODOM; PD000290; -; 1.  
 DR POSITE; PS00077; COX1; 1.  
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
 KW Respiratory chain; Transmembrane.  
 FT NON\_TER 353 353  
 SQ SEQUENCE 353 AA; 38214 MW; 3FA28DC6FFAF2FEA CRC64;

alignment\_scores:  
 Quality: 77.00 Length: 145  
 Ratio: 1.041 Gaps: 9  
 Percent Similarity: 51.034 Percent Identity: 28.276

alignment\_block:

US-09-540-234-1 x 047467 ..

Align seg 1/1 to: 047467 from: 1 to: 353

```

92 GTCGCGAGGCATATGGAACATCCCTGCCCTTTGAGGTCGCTC 141
|||||:|||||: ||| :||| ||| |||||:
109 AlaGlyThrGly...TrpThrValTyrProProLeuSerValVal 124
|||||:|||||: ||| :||| ||| |||||:
142 AAACATGCCCTCGAGAGCGTGGTGAAGCCGTAGTTGTACAAGACTAAC 191
|||||:|||||: ||| :||| ||| |||||:
124 aHisSer.....GlyProSerVal.AspLeuAlaIlePheSerLeuHis 138
|||||:|||||: ||| :||| ||| |||||:
192 AACGCCATGGCGTCGATGTTATCGAGGCCATCGTGTGCTGTTACTGT 241
|||||:|||||: ||| :||| ||| |||||:
139 LeuSerGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleThrIle 155
|||||:|||||: ||| :||| ||| |||||:
242 GTAGCGCTG.....ATGACTTCCTGCTGCTACTCCG 273
|||||:|||||: ||| :||| ||| |||||:
155 eGlnAsnMetArgAlaProGlyMetSerMetHisArgLeuProLeuPheV 172
|||||:|||||: ||| :||| ||| |||||:
274 CTTGAGGCCATTGTAGTGTGTTTACCGTGCAGACACTCGAAGACTCTCT 323
|||||:|||||: ||| :||| ||| |||||:
172 aITrp.....AlaIleTyrIleThrAlaPheLeuLeuLeu 184
|||||:|||||: ||| :||| ||| |||||:
324 GCT.....GTACTCGCGGAGCATAGTCTGTTGCTG..... 356
|||||:|||||: ||| :||| ||| |||||:
185 AlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAs 201
|||||:|||||: ||| :||| ||| |||||:
357 .....CCTGGTCTGGGACACCTGTGTACT 381
|||||:|||||: ||| :||| ||| |||||:
201 nPheAsnThrSerPheAspProAlaGlyGlyAspProIleLeu 218
|||||:|||||: ||| :||| ||| |||||:
382 ACAGTACTTCTGTTT.....CCTTTGTAGTGTACTACTT 416
|||||:|||||: ||| :||| ||| |||||:
218 yGlnHisLeuPheThrPhePheGlyHisProGluValTyrIleLeuIle 234
|||||:|||||: ||| :||| ||| |||||:
417 CMT.....AGCATGTACTC 431
|||||:|||||: ||| :||| ||| |||||:
235 LeuProGlyPheGlyIleIleSerHisIleLeu 245
|||||:|||||: ||| :||| ||| |||||:

```

seq\_name: sp\_plant:Q9LJ36

seq\_documentation\_block:

```

ID Q9LJ36 PRELIMINARY; PRT; 474 AA.
AC Q9LJ36;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0025005";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP001072; BAA89546.1; -;  
 DR InterPro; IPR001623; -;  
 DR Pfam; PF00226; DnaJ; 1.  
 DR PROSITE; PS00076; DnaJ\_2; 1.  
 DR SMART; SM00271; DnaJ; 1.  
 SQ SEQUENCE 474 AA; 51275 MW; 97E9CF9F70291705 CRC64;

alignment\_scores:

Quality: 77.00 Length: 128  
 Ratio: 1.013 Gaps: 5  
 Percent Similarity: 59.375 Percent Identity: 25.000

alignment\_block:

US-09-540-234-1 x Q9LJ36 ..

Align seg 1/1 to: Q9LJ36 from: 1 to: 474

```

2 GGTCCACGGGCCCTTATGTATGCAGGAGCGCCTACTGGGTCACTGAAGG 51
|||||:|||||: ||| :||| ||| |||||:
142 GlySerTyrIleLeuLeuPheAlaLeuGlyGlyHisGlyHisAl 158
|||||:|||||: ||| :||| ||| |||||:
52 AAAGGACAAAGACACCTCCAAGCCGACGCTACTGTAGCTGGCAGCG 101
|||||:|||||: ||| :||| ||| |||||:
158 aHisasnHisProMetGluLys.....MetAlaValAlaGlyLeuV 172
|||||:|||||: ||| :||| ||| |||||:
102 CATATGGAACACAGTCCTGCCCTTTGAGGTCGCTCTCAACATGCC 151
|||||:|||||: ||| :||| ||| |||||:
172 al.....LeuValProAlaLeuSerProCysAlaThrThrLeuPro 185
|||||:|||||: ||| :||| ||| |||||:
152 CTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAACACGCCAATG 201
|||||:|||||: ||| :||| ||| |||||:
186 Val.....PheLeu.AlaValGlyAsnSerSers 195
|||||:|||||: ||| :||| ||| |||||:
202 GCGTCGATGTTATCGAGCCATTCGTCGTGCTTACTGTGTAGCGCTTG 251
|||||:|||||: ||| :||| ||| |||||:
195 erMetMetIleLeuAlaIleIleValLeuLeuPheSerThrIleThrVal 211
|||||:|||||: ||| :||| ||| |||||:
252 ATGACTTCCTGCTGCTACTCCGCTTGGAGCCATTGAGTGTGTTTACCG 301
|||||:|||||: ||| :||| ||| |||||:
212 MetThrSerLeuValAlaLeuSerPheTyrGlyAlaSerGlnIleLysPh 228
|||||:|||||: ||| :||| ||| |||||:
302 TGCAGCACTCGAAGACTTCCTTGTGCTACTCGCTGGCAGCATAGTCTGTT 351
|||||:|||||: ||| :||| ||| |||||:
228 eHisTrpValGluArgTyrAspLysIleLeuValGlySerValLeuCysL 245
|||||:|||||: ||| :||| ||| |||||:
352 TCCTGCTGCTGCTGGGACACCTGTGTACTAC 383
|||||:|||||: ||| :||| ||| |||||:
245 euVal...GlyValLeuThrTyrValPheHis 254
|||||:|||||: ||| :||| ||| |||||:

```

seq\_name: sp\_organelle:P92800

seq\_documentation\_block:

```

ID P92800 PRELIMINARY; PRT; 528 AA.
AC P92800;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1.
OS Pyraliella littoralis.
OC Mitochondrion.
OC Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales;
OC Acinetosporaceae; Pyraliella.
OX NCBI_TaxID=2885;
RN [1]
RP SEQUENCE FROM N.A.
RA Fontaine J., Goux D., Kloareg B., Loiseaux-de Goer S.;

```



170 LeuPheValTrpSerValLeuLeuThrAlaPheLeuLeuLeuSer... 185  
 317 CTTCTTGTCTGCTACTCGCGGAGCATAGCTGTTTGTCTG... 356  
 186 ....LeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgA 201  
 357 .....CCTGGTCTGGGACACCTGTGTAC 380  
 201 snPheAsnThrThrPhePheAspProAlaGlyGlyAspProValLeu 217  
 381 TACAGTACTTCGTTT.....CCTTTGTAGTGGTACTACT 415  
 218 PheGlnHisLeuPheThrPhePheGlyHisProGluValTyriLeuIleu 234  
 416 TCTT.....AGCCATGACTCGGTGCT 437  
 234 eleuProGlyPheGlyIleValSerHisIleValSer 247  
 seq\_name: sp\_organelle:Q9TE64  
 seq\_documentation\_block:  
 ID Q9TE64 PRELIMINARY; PRT; 353 AA.  
 AC Q9TE64;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).  
 GN COXI.  
 OS Thalassiosira nordenskioldii.  
 OG Mitochondrion.  
 OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;  
 OC Thalassiosirophyceidae; Thalassiosirales; Thalassiosiraceae;  
 OC Thalassiosira.  
 OX NCBI\_TaxID=83372;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CCMP992;  
 RA Ehara M., Inagaki Y., Watanabe K.I., Ohama T.;  
 RT "Phylogenetic analysis of diatom cox1 genes and implications of  
 RT fluctuating GC content on mitochondrial genetic code evolution.";  
 RL Curr. Genet. 0:0-0(1999).  
 CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4  
 CC FERRICCYTOCHROME C (BY SIMILARITY).  
 CC -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).  
 CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 DR EMBL; AB020229; BAA96614.1; -;  
 DR InterPro; IPR000883; -;  
 DR Pfam; PF00115; COX1; 1.  
 DR PRINTS; PR01165; CYCOXIDASEI.  
 DR PROSITE; PS00077; COX1; 1.  
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
 KW Respiratory chain; Transmembrane.  
 FT NON\_TER 1 1  
 FT NON\_TER 353 353  
 SQ SEQUENCE 353 AA; 38158 MW; F107054A59A65521 CRC64;

alignment\_scores:  
 Quality: 76.50 Length: 130  
 Ratio: 1.142 Gaps: 5  
 Percent Similarity: 51.538 Percent Identity: 26.154  
 alignment\_block:  
 US-09-540-234-1 x Q9TE64 ..

Align seg 1/1 to: Q9TE64 from: 1 to: 353

135 TCGTCTCAACATGCCCTCGAGAGCGTGGTGAAGCCGTAGTTGTACAA 184  
 120 SerSerAlaIleHisSerGlySerValAspLeuAlaIlePheSe 136  
 185 GACTAACACGCCAATGGCGTCGATGTTATCGAGCCATTGCTGCTGT 234  
 136 rleuHisLeuSerGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleC 153  
 235 TTACGTGTAGCGCTTGAGACT.....TCCTTGCTG 266  
 153 ysThrIlePheAsnMetArgValLysSerLeuSerPheHisAsnLeuPro 169  
 267 TACTCCGCTTGGAGCCATTGAGTGTCTGTTTACCGTGCAGCACTCGAAGA 316  
 170 LeuPheValTrpSerValLeuIleThrAlaPheLeuLeuLeuSer... 185  
 317 CTTCTTGTCTGCTACTCGCTGGCAGCATAGCTGTTTGTCTG... 356  
 186 ....LeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgA 201  
 357 .....CCTGGTCTGGGACACCTGTGTAC 380  
 201 snPheAsnThrThrPhePheAspProAlaGlyGlyAspProValLeu 217  
 381 TACAGTACTTCGTTT.....CCTTTGTAGTGGTACTACT 415  
 218 PheGlnHisLeuPheThrPhePheGlyHisProGluValTyriLeuIle 234  
 416 TCTT.....AGCCATGACTCGGTGCT 437  
 234 eleuProGlyPheGlyIleValSerHisIleValSer 247  
 seq\_name: sp\_organelle:Q9G407  
 seq\_documentation\_block:  
 ID Q9G407 PRELIMINARY; PRT; 511 AA.  
 AC Q9G407;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE CYTOCHROME C OXIDASE SUBUNIT 1.  
 GN COX1.  
 OS Lithobius forficatus.  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Myriapoda; Chilopoda;  
 OC Pleurostigmophora; Lithobiomorpha; Lithobiidae; Lithobius.  
 OX NCBI\_TaxID=7552;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20558576; PubMed-11095730;  
 RA Lavrov D.V., Brown W.M., Boore J.L.;  
 RT "A novel type of RNA editing occurs in the mitochondrial tRNAs of the  
 RT centipede Lithobius forficatus";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:13738-13742(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Lavrov D.V., Brown W.M., Boore J.L.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF309492; AAG39986.1; -;  
 KW Mitochondrion.  
 SQ SEQUENCE 511 AA; 56589 MW; E01618C6ED6956AA CRC64;

alignment\_scores:  
 Quality: 76.50 Length: 159  
 Ratio: 0.981 Gaps: 9  
 Percent Similarity: 49.057 Percent Identity: 28.302  
 alignment\_block:  
 US-09-540-234-1 x Q9G407 ..





OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlmann P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
RA Davidson S., Wohlmann P., Bauer C.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX Waterston R.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF02976; AAC69087.1; -;  
DR InterPro; IPR000276; -;  
SQ SEQUENCE 349 AA; 39882 MW; 43EC5A1DB130CBDD CRC64;

alignment\_scores:  
Quality: 76.00 Length: 108  
Ratio: 1.070 Gaps: 9  
Percent Similarity: 65.741 Percent Identity: 32.407  
alignment\_block:  
US-09-540-234-1/rev x O16955 ..  
Align seg 1/1 to: O16955 from: 1 to: 349

305 TCCAGCGTAAACAGCAGCTCAATGCTCCAGCGGAGTACAGCAGGAAG 256  
181 CysThrThrGluSerThrArg...ArgProSerAlaAlaLysSerVally 196  
255 TCATCAAGCGGTACACAGTAAACAGCAGCAGCAATGGCTCCGATACATCG 206  
196 slyArgCysLeu.ThrValAspAsnPhcGlnPheLeuProSerLeuAla 212  
205 AGGCCA...TTGGCGTGTGTAGCTTTGTACAACTACGGCTTCACCGCAG 159  
213 LeuProIleLeuThrPheLeuLeuLeuArgGluMetArgIleLeuLysG1 229  
158 TCTCCAGGGCATGTTTTCAGACGACCCCTCAAAGGCGGACGACTTTTC 109  
229 n.SerArg.....SerThrLeuGlnArgSerGlyValVal 241  
108 .....CATATGCGGTGTCAGCTACAGTACGGTGTCCGGGCT 71  
242 AspAsnGluGluLysTyrglyLeuPro...ThrLysLeuIleAlaLeuLe 257  
70 GGAGGTGTTTGTCTTCCTTCAGTACGCCACTAGCC..... 31  
257 urThrIleThr...SerPheLeuAlaAspThrProLeuGlyCysIleGlyI 273  
30 ..GCTCCTGCATACATAGG 13  
273 leAlaLysAlaPheIleGln 279

seq\_name: sp\_human:Q9P234  
seq\_documentation\_block:  
ID Q9P234 PRELIMINARY; PRT; 638 AA.  
AC Q9P234;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE KIAA1494 PROTEIN (FRAGMENT).  
GN KIAA1494.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20277482; PubMed=10819331;  
RA Nagase T., Kikuno R., Ishikawa K., Hirokawa M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human  
RT genes XVII. The complete sequences of 100 new cDNA clones from brain  
RT which code for large proteins in vitro.";  
RL DNA Res. 7:143-150(2000).  
DR EMBL; AB040927; BAA96018.1; -;  
DR InterPro; IPR001452; -;  
DR Pfam; PF00018; SH3; 2.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR PROSITE; PS50002; SH3; 2.  
FT NON\_TER 1  
SQ SEQUENCE 638 AA; 65516 MW; 01650B629ACD3053 CRC64;  
alignment\_scores:  
Quality: 76.00 Length: 113  
Ratio: 1.310 Gaps: 5  
Percent Similarity: 51.327 Percent Identity: 29.204  
alignment\_block:  
US-09-540-234-1/rev x Q9P234 ..  
Align seg 1/1 to: Q9P234 from: 1 to: 638  
398 GGGAAACGAAGTACTGTAGTACACAGGTGTCCCGACAGCAGCAGCAAC 349  
183 GlySerThrAspGlnIleAlaHisLeuArgProGlnThrArgProSerVa 199  
348 AGACTATGCTGCCAGCGAGTACAGC.....AGGAA..... 318  
199 ltyrValAlaIleTyProTyThrProArgLysGluAspGluLeuGluL 216  
317 .....GTCTTCGAGTGTCTGCACG..... 300  
216 euArgLysGlyGluMetPheLeuValPheGluArgCysGlnAspGlyTrp 232  
299 GTAAACAGCAGCTCAATGGCTCCCAAGC.....GGAGTACAGCAAGGAAG 256  
233 PheLysGlyThrSerMetHisThrSerLysIleGlyValPheProGlyAs 249  
255 TCATCAAGCGGTACACAGTAAACAGCAGCAGCAATGGCTCCGATACATCG 206  
249 ntyrValAlaProValThrArgAlaValThrAsnAlaSerGln..... 263  
205 ACGCCATTGGGGTGTGTAGTCTTGTACAAACTACGGCTTCACCGAGTCT 156  
264 .....AlaLysValProMetSerThrAlaGlyGlnThrSer 275  
155 CGAGGGGCGATGTTTGAGACGACCCCTCAAAGGCGGAGGA 117  
276 ArgGlyValThrMetValSerProSerThrAlaGlyGly 288  
seq\_name: sp\_human:Q9HAM2  
seq\_documentation\_block:





THE UNIVERSITY OF CHICAGO PRESS



```
alignment_scores:
  Quality: 74.50
  Ratio: 0.776
  Percent Similarity: 50.000
  Length: 192
  Gaps: 9
  Percent Identity: 23.958

alignment_block:
```

```

US-09-540-234-1/rev x 017047  ..
Align seg 1/1 to: 017047 from: 1 to: 707

442 CGACAGACAGCAGTACATGGCTGCAAGTAGT.....AC 408
||| ::| ||||| |||:::|||||
9 ArgGlnAsnProSerThrGluLeuTyrLysSerThrArgAlaThrThrTh 25
|||:::|||||:::|||||:::|||||:::|||||:::|||||
407 CACTACAAAGGAAAGCAAGTAGTCTGTAGTACACAGGTGCTCCACAGACCAG 358
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
25 rThrThrSerSerSerGlnArgAsnTyrAsn.....A 37
357 GCAGCAACAGCAGTATGCTGCCAGCAGTACAGCAAGGAGTCTTCGAGT 308
:::|||||:::|||||:::|||||:::|||||:::|||||
37 sPAsnAsnGlnAsnThrSerSerSerSerGlyAsnLysLysGluSerSer 53
307 GCTGCACGTAACACAGCACTCAAT...GCCTCCAGCGGAGTACAGCAA 261
:::|||||:::|||||:::|||||:::|||||:::|||||
54 SerSer.SerLysGlnHisSerSerLysLysSerLysLysSerAsnSerL 70
:::|||||:::|||||:::|||||:::|||||:::|||||
260 GGAAGTCATCAAGCGCTACA.....CAGTAAACA..... 232
:::|||||:::|||||:::|||||:::|||||:::|||||
70 yLysAsnArgSerProSerProGlnProGlnLeuThrIleLysSerAla 86
:::|||||:::|||||:::|||||:::|||||:::|||||
231 .....GCACAGCAATGGCTCCGA.....TA 212
:::|||||:::|||||:::|||||:::|||||
87 IleLeuIleGlnLysTrpTyrArgArgCysGluAlaArgLeuGluAlaAr 103
211 ACATCGACCCCATTTGGCGTTGTAGTCTTGTACAAACTACGGC..... 169
:::|||||:::|||||:::|||||:::|||||:::|||||
103 gArgArgAlaThrTrpGlnIlePheThrAlaLeuGluTyrAlaGlyGlu 120
168 .....TTCCACCGAGCTCTCGAGGGGCAG 145
:::|||||:::|||||:::|||||:::|||||:::|||||
120 lnaSpGlnLeuLysLeuTyrAspPheAlaAspValIleArgAlaMet 136
144 TTTGAGACGACCCCTCAAAAGGCA..... 121
||| ::| ||| ::|
137 AlaGluAsnGlyLysGlyGlyValGluAsnGlyArgAsnSerProLe 153
120 ...GGGACTGTTTCCATTATGCGTGCAGCTACAGTAGCCGTGTCGG 75
:::|||||:::|||||:::|||||:::|||||:::|||||
153 uMetSerAlaLeuSerHisTyrAlaLysProSerLeuMetAspSerGlu 170
74 GCTTGGAGGTGCTTTGTCTTCTTCAGATGACCCACTAGCGCTCT 25
|||:::|||||:::|||||:::|||||:::|||||
170 lylGluThrValLysLysMetLeuGluAspThrSerProThrAsnValasp 186
24 GCATACATAAGGCCCGCTGGACCC 1
::: |||:::|||||
187 lIeAspArgAsnTyrLysGlyPro 194
seq_name: sp_organelle:O99383

```

```

seq_documentation_block:
ID O99383 PRELIMINARY; PRT; 353 AA.
AC O99383;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COXI.
OS Ectocarpus sp.
OC Mitochondrion.
OC Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales; Ectocarpaceae;
OC Ectocarpus.
OX NCBI_TaxID=74372;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN-HIROSHIMA, JAPAN;
RA Ehara M., Inagaki Y., Watanabe K.I., Hayashi-Ishimaru Y., Kawai H.,
RA Ohama T.;
RT *Phylogenetic analyses of heterokont algae with emphasis on the AUA

```

```

RT codon reassignment in mitochondria.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4
CC FERRICYTOCHROME C.
CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AF037994; AAC94980.1; -.
DR InterPro; IPR000883; -.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PR01165; CYCOXIDASE1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS00077; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT NON_TER 1
FT NON_TER 353
SQ SEQUENCE 353 AA; 38107 MW; 51E80DAB2B3A01A3 CRC64;

```

```

alignment_scores:
Quality: 74.00 Length: 143
Ratio: 1.014 Gaps: 8
Percent Similarity: 51.049 Percent Identity: 26.573

alignment_block:
US-09-540-234-1 x O99383  ..

```

```

Align seg 1/1 to: O99383 from: 1 to: 353

92 GTGGCAGCGCATATGGAACAGTCCTGCCCTTTGAGGGTGGCTCTC 141
109 AlaGlyThrGly...TrpThrValTyrProProLeuSerGlyIleGlnAl 124
142 AAACATGCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAAGACTAAC 191
124 aHisSer.....GlyProSerVal.AspLeuAlaIlePheSerLeuHis 138
192 AACGCCAATGGCGTCGATTTATCGGAGCCATTCGGTGCCTGTTACTGT 241
139 LeuSerGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleThrThrI 155
242 GTAGCGCTTG.....ATGACTTCCTTGTGCTACTCCG 273
155 ePheAsnMetArgAlaProGlyMetThrMetAspArgValProLeuPheV 172
274 CTTGGAGCCATTGTAGTGTCTTTTACCGTCGACGACTCGAAGACTTCCTT 323
172 aITrpSerValLeuIleThrAlaPheLeuLeuLeuSer.....Leu 186
324 GCTGTACTCGTGGCAGCATAGTCTGTTTGTGCTG..... 356
187 ProValLeuAlaGlyGlyIleThrMetLeuLeuThrAspArgAsnLeuAs 203
357 .....CCTGGTCTGGGACACCTGTGTACTACTACAGTA 387
203 nThrThrPhePheAspProAlaGlyGlyAspProValLeuTyrGlnH 220
388 CTTGCTTT.....CCTTTGTAGTGGTACTACTTCTT... 419
220 lLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 236
420 .....AGCCATGTACTC 431
|||||:::|||||

```



245 GCGCTTGATGACTTCCTTGTG  
          :::|  
          :::|  
          :::|



SEQUENCE 485 AA; 53654 MW; 48DB4A9A145678E8 CRC64;

260 ValIleVal 262

**THIS PAGE BLANK (USPTO)**

OM of: US-09-540-234-1 to: SwissProt\_39:\* out\_format : pfs

Date: Sep 7, 2001 5:07 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+nt2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09540234/runat_07092001_145856_19530/app_query.fasta_1.507
-DB=SwissProt_39 -OFMT=fastan -SUFFIX=rs -GAPOP=12.000
-CAPEXT=4.000 -MINMATCH=0.100 -LOOCL=0.000 -LOOEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09540234 -CGN1_1_24 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

Query: US-09-540-234-1

Query length: 444

Database: SwissProt\_39:\*

Database sequences: 93435

Database length: 34255486

Search time (sec): 36.670000

score\_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
SwissProt_39:NTPR_RAT +	81.50	136.57	0.8763		661	P28573 rattus norvegicus (rat)
SwissProt_39:COX1_ALBICO +	79.00	134.00	1.58		509	P48887 albinaria coerulea (l)
SwissProt_39:LEUK_MOUSE -	77.00	132.32	2.53		395	P15702 mus musculus (mouse)
SwissProt_39:NKCR_HUMAN -	77.00	121.12	2.88		1462	P30414 homo sapiens (human)
SwissProt_39:NTPR_HUMAN +	75.00	124.39	4.35		636	P09884 homo sapiens (human)
SwissProt_39:AGSR_BOVIN +	74.00	135.79	4.78		134	P56413 bos taurus (bovine)
SwissProt_39:2FH2_DROME -	73.50	108.22	7.32		3005	P28167 drosophila melanogaster
SwissProt_39:MEYK_BACSU +	73.00	124.51	6.81		400	P54419 bacillus subtilis, s-
SwissProt_39:COX1_LUMTE +	73.00	122.38	6.97		513	P34941 lumbricus terrestris
SwissProt_39:COX1_PROWI +	72.50	121.40	7.89		514	P05143 prototheca wickerhamii
SwissProt_39:COX1_RHILE +	72.50	121.01	7.93		538	P08855 rhizobium leguminosar
SwissProt_39:ESCL_SCHPO -	72.00	122.31	8.74		413	P04635 schizosaccharomyces
SwissProt_39:CCAD_CHICK -	71.50	107.07	11.63		2190	P07370 gallus gallus (chick)
SwissProt_39:HRX_HUMAN -	71.50	101.99	12.32		3969	P03164 homo sapiens (human)
SwissProt_39:SPV2_MOUSE -	70.50	121.74	12.33		315	P09xv8 mus musculus (mouse)
SwissProt_39:MR15_DROME -	70.00	118.23	14.36		424	P09011 drosophila melanogaster
SwissProt_39:GCR2_YEAST -	69.50	115.30	16.61		534	P01722 saccharomyces cerevis
SwissProt_39:FIMC_BORPE -	69.00	110.13	19.71		873	P33410 bordetella pertussis
SwissProt_39:M84D_DROME +	68.50	131.01	17.40		68	P01845 drosophila melanogaster
SwissProt_39:TBAL_PELFA +	68.00	113.82	23.67		459	P34031 pelvetia fastigiata
SwissProt_39:YK25_CAEEL -	68.00	109.07	24.99		789	P33332 caenorhabditis elegans
SwissProt_39:COX1_ACACA +	68.00	108.20	25.24		873	P03730 acanthamoeba castellan
SwissProt_39:COX1_PHYME +	67.00	111.19	30.55		492	P02211 phytophthora megasper
SwissProt_39:COX1_LOCHI +	67.00	110.84	30.67		512	P03421 locusta migratoria (m
SwissProt_39:COX1_RHISA +	67.00	110.84	30.67		512	P09818 rhipicephalus sanguin
SwissProt_39:CIK3_MOUSE -	67.00	110.58	30.76		528	P16390 mus musculus (mouse)
SwissProt_39:MBHL_ALCHI +	67.00	109.19	31.25		621	P33374 alcaligenes hydrogenc
SwissProt_39:FABH_HAEIN +	66.50	111.87	33.92		406	P43710 haemophilus influenza
SwissProt_39:FRA_DROME -	66.50	108.60	35.21		595	P21525 drosophila melanogaster
SwissProt_39:PRIA_BACSU -	66.50	106.01	36.26		805	P94461 bacillus subtilis, pr
SwissProt_39:AGSR_HUMAN +	66.00	120.52	34.40		132	P00253 homo sapiens (human)
SwissProt_39:GBA5_DICDI +	66.00	112.25	37.80		347	P34043 dictyostelium discoide
SwissProt_39:COX1_PELSI +	66.00	108.89	39.27		514	P07967 pelomedusa subrufa (a
SwissProt_39:COX1_MARPO +	66.00	108.75	39.33		522	P26856 marchantia polymorpha
SwissProt_39:COX1_BETVU +	66.00	108.72	39.34		524	P24794 beta vulgaris (sugar
SwissProt_39:COX1_ORISA +	66.00	108.72	39.34		524	P14578 oryza sativa (rice)
SwissProt_39:COX1_ARATH +	66.00	108.67	39.37		527	P07063 arabidopsis thaliana
SwissProt_39:COX1_OENBE +	66.00	108.67	39.37		527	P08743 onocrotalaria bertiana
SwissProt_39:COX1_NAIZE +	66.00	108.66	39.37		528	P08742 zea mays (maize)
SwissProt_39:COX1_SORBI +	66.00	108.62	39.39		530	P05502 sorghum bicolor (sorg

```
SwissProt_39:SYR_TREPA + 66.00 107.72 39.79 589 ! 083803 treponema pallidum.
SwissProt_39:ZG20_XENLA - 66.00 106.55 40.33 675 ! P18714 xenopus laevis (af
SwissProt_39:MUC1_RAT + 66.00 104.71 41.18 837 ! P98089 rattus norvegicus (r
SwissProt_39:CR2_MOUSE - 66.00 102.98 42.00 1025 ! P19070 mus musculus (mous
SwissProt_39:MAXX_HUMAN + 66.00 102.08 42.43 1139 ! P49641 homo sapiens (hum
```

seq\_name: SwissProt\_39:NTPR\_RAT

seq\_documentation\_block:

```
ID NTPR_RAT STANDARD; PRT; 661 AA.
AC P28573;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DE SODIUM-DEPENDENT PROLINE TRANSPORTER.
GN SLC6A7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=92265305; PubMed=1350201;
RA Fremieu R.T., Jr., Caron M.G., Blakely R.D.;
RT "Molecular cloning and expression of a high affinity L-proline
RT transporter expressed in putative glutamatergic pathways of rat
RT brain.";
RL Neuron 8:915-926(1992).
CC -! FUNCTION: TERMINATES THE ACTION OF PROLINE BY ITS HIGH AFFINITY
CC SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -! TISSUE SPECIFICITY: EXPRESSED IN SUBPOPULATIONS OF PUTATIVE
CC GLUTAMATERGIC PATHWAYS OF RAT BRAIN.
CC -! SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNP).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
-----
EMBL; M88111; AAA41541.1; -.
InterPro; IPR000175; -.
Pfam; PF00209; SNF; 1.
PRINTS; PR00176; NANEUSMPORT.
PROSITE; PS00610; NA_NEUROTRAN_SYM_1; 1.
PROSITE; PS00754; NA_NEUROTRAN_SYM_2; 1.
PROSITE; PS0267; NA_NEUROTRAN_SYM_3; 1.
KW Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
KW Symport; Amino-acid transport.
FT DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 46 66 1 (POTENTIAL).
FT TRANSMEM 74 93 2 (POTENTIAL).
FT TRANSMEM 117 137 3 (POTENTIAL).
FT DOMAIN 138 214 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 215 233 4 (POTENTIAL).
FT TRANSMEM 242 259 5 (POTENTIAL).
FT TRANSMEM 295 312 6 (POTENTIAL).
FT TRANSMEM 324 345 7 (POTENTIAL).
FT TRANSMEM 378 397 8 (POTENTIAL).
FT TRANSMEM 425 443 9 (POTENTIAL).
FT TRANSMEM 459 479 10 (POTENTIAL).
FT TRANSMEM 500 519 11 (POTENTIAL).
FT TRANSMEM 538 556 12 (POTENTIAL).
FT DOMAIN 557 661 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 661 AA; 73684 MW; 8154F26C55A72D4A CRC64;
```

```

alignment_scores:
  Quality: 81.50      Length: 105
  Ratio: 1.455        Gaps: 2
  Percent Similarity: 53.333  Percent Identity: 24.762

alignment_block:
US-09-540-234-1 x NTPR_RAT ..
Align seg 1/1 to: NTPR_RAT from: 1 to: 661

105 AATGGAACACAGTCCTCCCTTTGAGGCTGCTCAACATGCCCCCTC 154
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
174 AspGlyAsnGlyAlaLeuProLeuAsnLeuSerThrValSerProse 190
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
155 GAGACGTCGCGAGCCGAGTGTGTACAGACTACACAGCCATGGCG 204
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
190 rGluGluTrpSerArgTrpValLeuHisIleGlnGlySerGlnGly 207
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
205 TCGATGTTATCGGAGCCATCGTGTCTGTTACTGTGTAGCGCTGATG 254
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
207 leGlyArgProGlyGluIleArg..... 214
255 ACTTCCTCTGCTACTCCGCTTGGAGCCATTGAGTGTGTGTACCGTC 304
    ::::: ::::: ::::: ::::: ::::: :::::
215 ..... 224
305 AGCACTCGAAGACTTCCTGTGCTGCTACTCGCTGGCAGCAGTAGTGT 354
    ::::: ::::: ::::: ::::: ::::: :::::
224 atpValIleValPheLeuCysIleLeuLysGly.....V 236
355 TGCGTGTGCTGGGACACCTGTGTACTACAGTACTTCGTTTCCCTTGT 404
    ::::: ::::: ::::: ::::: ::::: :::::
236 alySerSerGlyLysValValTrpPheThrAlaThrPheProTrpLeu 252
    ::::: ::::: ::::: ::::: ::::: :::::
405 GTGGTACTACTTCTT 419
    ::::: ::::: ::::: ::::: ::::: :::::
253 IleLeuLeuMetLeu 257

seq_name: SwissProt_39:COX1_ALBCO
seq_documentation_block:
ID COX1_ALBCO STANDARD; PRT; 509 AA.
AC P48887;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COI.
OS Albinaria coerulea (Land snail).
OG Mitochondrion.
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OX Clausiliidae; Alopinae; Albinaria.
OX NCBI_TaxID=42349;
RN [1]
RP SEQUENCE FROM N.A. PubMed-7498775;
RX MEDLINE=96120351;
RA Hatzoglou E., Rodakis G.C., Lecanidou R.;
RT "Complete sequence and gene organization of the mitochondrial genome
of the land snail Albinaria coerulea.";
RL Genetics 140:1353-1366(1995).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
AND COPPER B.
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
4 FERRICYTOCHROME C.
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

```

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

CC DR EMBL; X83390; CAA58307.1; -;  
 CC DR HSSP; P00396; 10CC.  
 CC DR InterPro; IPR000883; -;  
 CC DR InterPro; IPR002428; -;  
 CC DR Pfam; PF00115; COX1; 1;  
 CC DR PRINTS; PRO1165; CYCOXIDASE1.  
 CC DR PROSITE; PS00077; COX1; 1.  
 CC KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
 CC FT Respiratory chain; Inner membrane.  
 CC FT METAL 57 57 IRON (HEME A) (PROBABLE).  
 CC FT METAL 236 236 COPPER B (PROBABLE).  
 CC FT METAL 240 240 COPPER B (PROBABLE).  
 CC FT METAL 285 285 COPPER B (PROBABLE).  
 CC FT METAL 286 286 COPPER B (PROBABLE).  
 CC FT METAL 371 371 IRON (HEME A3) (PROBABLE).  
 CC FT METAL 373 373 IRON (HEME A) (PROBABLE).  
 CC SQ SEQUENCE 509 AA; 57034 MW; 07236D1343151A6B CRC64;

alignment\_scores:  
 Quality: 79.00 Length: 143  
 Ratio: 1.097 Gaps: 8  
 Percent Similarity: 50.350 Percent Identity: 27.972

alignment\_block:  
US-09-540-234-1 x COX1\_ALBCO ..  
Align seg 1/1 to: COX1\_ALBCO from: 1 to: 509

92 GCTGGCAGCGCAATATGGAACAGTCCTCCCTTTGAGGCTGCTCTC 141  
 ::::: ::::: ::::: ::::: ::::: ::::: :::::  
 118 AlaGlyThrGly...TrpThrValTrpPro.ProLeu.....SerSerS 131  
 142 AACATGCCCTCGAGACGCTGCGTGAAGCCGTAGTTGTACAGACTAAC 191  
 ::::: ::::: ::::: ::::: ::::: ::::: :::::  
 131 erLeuAlaHisSerGlyAlaSerValAspLeuAlaIlePheSerLeuHis 147  
 192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGCTGTGTTACTGT 241  
 ::::: ::::: ::::: ::::: ::::: ::::: :::::  
 148 LeuAlaGlyMetSerSerIleLeuGlyAlaIleAsnPheIleThrIle 164  
 242 GTAGCGCTTG.....ATGACTTCCTGCTGCTACTCCG 273  
 ::::: ::::: ::::: ::::: ::::: ::::: :::::  
 164 epheAsnMetArgSerProGlyMetThrMetGluArgValSerLeuPheV 181  
 274 CTTGGAGCCATTGAGTGTGTTTACCGTGCACCACTCGAAGACTTCCTT 323  
 ::::: ::::: ::::: ::::: ::::: ::::: :::::  
 181 alTrpSerIleLeuValThrValPheLeuLeuLeuSer.....Leu 195  
 324 GCTGTACTCGCTGGCAGCATAGTCTGTTGCTG..... 356  
 ::::: ::::: ::::: ::::: ::::: ::::: :::::  
 196 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 212  
 357 .....CCTGGTGGGACACCTGTGTACTACAGTA 387  
 212 nThrSerPhePheAspProAlaGlyGlyGlyAspProIleLeuTrpGlnH 229  
 388 CTCGTTT.....CCCTTTGTAGTGTACTTCTT... 419  
 ::::: ::::: ::::: ::::: ::::: ::::: :::::  
 229 isLeuPheTrpPhePheGlyHisProGluValTrpIleLeuLeuLeuPro 245  
 420 .....AGCCATGTACTC 431  
 246 GlyPheGlyMetIleSerHisIleLeu 254

```

seq_documentation_block:
ID      LEUK_MOUSE      STANDARD;      PRT;      395 AA.
AC      F15702;
DT      01-APR-1990 (Rel. 14, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      LEUKOSIALIN PRECURSOR (LEUCOCYTE STIALOGLYCOPROTEIN) (STALOPHORIN)
DE      (CD43 ANTIGEN) (LY 48) (B CELL DIFFERENTIATION ANTIGEN LP-3).
GN      SPN.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
[1]
RN      SEQUENCE FROM N.A.
RP      STRAIN=DBA/2J; TISSUE=Liver;
RC      MEDLINE=90269342; PubMed=2347365;
RX      Cyster J.G., Somoza C., Killeen N., Williams A.F.;
RY      "Protein sequence and gene structure for mouse leukosialin (CD43), a
RT      T lymphocyte mucin without introns in the coding sequence.";
RN      Eur. J. Immunol. 20:875-881(1990).
[2]
RN      SEQUENCE FROM N.A.
RP      STRAIN=B10.P; TISSUE=Liver;
RC      MEDLINE=90370495; PubMed=2144340;
RX      Dorfman K.S., Litaker K.S., Baecher C.M., Frelinger J.G.;
RY      "The nucleotide sequence of Ly 48 (mouse leukosialin, sialophorin):
RT      the mouse homolog of CD43.";
RN      Nucleic Acids Res. 18:4932-4932(1990).
[3]
RN      SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP      MEDLINE=94236703; PubMed=7514104;
RX      Shiota J., Nishimura H., Okamoto H., Yu B., Hattori S., Abe M.,
RA      Okada T., Nozawa S., Tsurui H., Hirose S.;
RY      "A unique murine CD43 epitope Lp-3: distinct distribution from
RT      another CD43 epitope S7";
RN      Cell. Immunol. 155:402-413(1994).
[4]
RN      SEQUENCE OF 345-383 FROM N.A.
RP      STRAIN=C57BL/6J;
RC      MEDLINE=90316596; PubMed=1973410;
RX      Baecher C.M., Dorfman K.S., Mattei M.-G., Frelinger J.G.;
RY      "CDNA cloning and localization of the mouse leukosialin gene (Ly48)
RT      to chromosome 7.";
RN      Immunogenetics 31:307-314(1990).
-1- FUNCTION: ONE OF THE MAJOR GLYCOPROTEINS OF THYMOCYTES AND T
    LYMPHOCYTES. PLAYS A ROLE IN THE PHYSICOCHEMICAL PROPERTIES OF
    THE T-CELL SURFACE AND IN LECTIN BINDING. PRESENTS CARBOHYDRATE
    LIGANDS TO SELECTINS. HAS AN EXTENDED RODLIKE STRUCTURE THAT COULD
    PROTRUDE ABOVE THE GYCOALYX OF THE CELL AND ALLOW MULTIPLE
    GLYCAN CHAINS TO BE ACCESSIBLE FOR BINDING.
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
-1- TISSUE SPECIFICITY: CELL SURFACE OF THYMOCYTES, T LYMPHOCYTES,
    NEUTROPHILS, PLASMA CELLS AND MYELOMAS.
-1- PTM: HAS A HIGH CONTENT OF SIALIC ACID AND O-LINKED CARBOHYDRATE
    STRUCTURES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; X17018; CAA34884.1; -
EMBL; X52609; CAA36840.1; -
EMBL; S70677; AAB30765.1; -
EMBL; M30693; AAB39457.1; -
PIR; S08065; S08065.

```





-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 -----



DR EMBL; Z99119; CAB15033.1; -  
 DR HSP; P04384; 1XRC.  
 DR Subtilist; BG11840; metK.  
 DR InterPro: IPR0021133; -  
 DR Pfam; PF00438; S-AdoMet\_synt; 1.  
 DR PROSITE; PS00376; ADOMET\_SYNTHETASE\_1; 1.  
 DR PROSITE; PS00377; ADOMET\_SYNTHETASE\_2; 1.  
 KW Transferase; One-carbon metabolism; ATP-binding.  
 FT NP\_BIND 129 134 ATP (POTENTIAL).  
 FT CONFLICT 26 26 S -> Y (IN REF. 1).  
 FT CONFLICT 111 111 Q -> P (IN REF. 1).  
 FT CONFLICT 140 140 A -> V (IN REF. 1).  
 SQ SEQUENCE 400 AA; 44042 MW; EF3D97DCD375714A CRC64;

alignment\_scores:  
 Quality: 73.00 Length: 57  
 Ratio: 1.825 Gaps: 0  
 Percent Similarity: 70.175 Percent Identity: 26.316

alignment\_block:

US-09-540-234-1/rev x METK\_BACSU ..

Align seg 1/1 to: METK\_BACSU from: 1 to: 400

365 CAGACAGGAGCAGCAACAGACTATGCTGCCAGCGAGTACGACGAGGAGT 316  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 73 GlnThrIleLysGluIleGlyTyrThrArgAlaLysTyrGlyPheAspAl 89  
 315 CTTGCGAGTGTGCGGTAACAGCACTCAATGGCTCCAGCGGAGTAC 266  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 89 aGluThrCysAlaValLeuThrSerIleAspGluGlnSerAlaAspIleA 106  
 265 AGCAAGGAGTCAATCAACGCGTACACAGTAAACAGCAGCAATGGCTCC 216  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 106 laMetGlyValAspGlnAlaLeuGluAlaArgGluGlyThrMetSerAsp 122  
 215 GATAACATCGACGCCATGGC 195  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 123 GluGluIleGluAlaIleGly 129

seq\_name: SwissProt\_39:COX1\_LUMTE

seq\_documentation\_block:

ID COX1\_LUMTE STANDARD; PRT; 513 AA.  
 AC Q34941;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 GN COI.  
 OS Lumbricus terrestris (Common earthworm).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;  
 OC Lumbricina; Lumbricidae; Lumbricus.  
 OX NCBI\_TaxId=6398;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96042914; PubMed=8536978;  
 RA Boore J.L., Brown W.M.;  
 RT "Complete sequence of the mitochondrial DNA of the annelid worm Lumbricus terrestris.";  
 RL Genetics 141:305-319(1995).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B.  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) - 2 H(2)O + 4 FERRICYTOCHROME C.  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

DR EMBL; U24570; AAC46864.1; -  
 DR HSP; P98002; 1AR1.  
 DR InterPro: IPR000883; -  
 DR InterPro: IPR002428; -  
 DR Pfam; PF00115; COX1; 1.  
 DR PRINTS; PR01165; CYCOXIDASE1.  
 DR PROSITE; PS00077; COX1; 1.  
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
 KW Respiratory chain; Inner membrane.  
 FT METAL 58  
 FT METAL 237 237 IRON (HEME A) (PROBABLE).  
 FT METAL 241 241 COPPER B (PROBABLE).  
 FT METAL 287 287 COPPER B (PROBABLE).  
 FT METAL 288 288 COPPER B (PROBABLE).  
 FT METAL 373 373 IRON (HEME A3) (PROBABLE).  
 FT METAL 375 375 IRON (HEME A) (PROBABLE).  
 SQ SEQUENCE 513 AA; 56808 MW; EC8573C300326B70 CRC64;

alignment\_scores:

Quality: 73.00 Length: 159  
 Ratio: 1.028 Gaps: 7  
 Percent Similarity: 44.654 Percent Identity: 24.528

alignment\_block:

US-09-540-234-1 x COX1\_LUMTE ..

Align seg 1/1 to: COX1\_LUMTE from: 1 to: 513

107 TGGAAACAGTCCTGCGCTTTTGGGGTCTCTCAACATGCCCTCGA 156  
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 100 TrpLeuLeuProSerLeuIleLeuValSerSerAlaAlaValG1 116  
 157 G.....  
 116 uLysGlyAlaGlyThrGlyTrpThrValTyrProLeuAlaSerAsnL 133  
 158 .....ACGTGCGTGAAGCGGTAGTTTGTACAGACTAACAA 193  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 133 euAlaHisAlaGlyProSerVal.AspleuAlaIlePheSerLeuHisLe 149  
 194 CGCAATGGCTCGATGTTATCGGAGCATTCGTCGTCTGTACTGTGT 243  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 149 uAlaGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleThrValI 166  
 244 AGCGCTTGATGACTTCCTTGTCTG.....TACTCGCT 275  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 166 leAsnMetArgTrpSerGlyLeuArgIleProLeuPheVal 182  
 276 TGGAGCCATTTGAGTGTGTTTACCGTCGACAGACTTCCTTGC 325  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 183 TrpAlaValLeuIleThrValValLeuLeuLeuSer.....LeuPr 197  
 326 TGTACTCGTCGACAGTACTCTTTTGTCTG..... 356  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 197 oValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnLeuAsnT 214  
 357 .....CCTGGTCTGGGACACCTGTGTACTACTACTACT 389  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 214 hrSerPhePheAspProAlaGlyGlyAspProIleLeuTyrGlnHis 230





[illegible]



RA Djabali M., Selleri L., Parry P., Bower M., Young B.D., Evans G.A.;  
 RT "A trithorax-like gene is interrupted by chromosome 11q23  
 translocations in acute leukaemias.";  
 RL Nat. Genet. 2:113-118(1992).  
 RN [5]  
 RP SEQUENCE OF 1251-1538 FROM N.A.  
 RX MEDLINE=94215165; PubMed=8162575;  
 RA Gu Y., Alder H., Nakamura T., Schichman S.A., Prasad R., Canaanani O.,  
 RA Saito H., Croce C.M., Canaanani E.;  
 RT "Sequence analysis of the breakpoint cluster region in the ALL-1 gene  
 involved in acute leukemia.";  
 RL Cancer Res. 54:2326-2330(1994).  
 RN [6]  
 RP SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).  
 RX MEDLINE=95322025; PubMed=7598802;  
 RA Mbangkollo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,  
 RA Rowley J.D., Diaz M.O.;  
 RT "The human MLL gene: nucleotide sequence, homology to the Drosophila  
 trx zinc-finger domain, and alternative splicing.";  
 RL DNA Cell Biol. 14:475-483(1995).  
 RN [7]  
 RP SEQUENCE OF 1212-1603 FROM N.A.  
 RX MEDLINE=95315013; PubMed=7794749;  
 RA Marchalek R., Grell J., Lochner K., Nilsson I., Siegler G.,  
 RA Zwickbronner I., Beck J.D., Fey G.H.;  
 RT "Molecular analysis of the chromosomal breakpoint and fusion  
 transcripts in the acute lymphoblastic SEM cell line with chromosomal  
 translocation t(4;11).";  
 RL Br. J. Haematol. 90:308-320(1995).  
 RN [8]  
 RP SEQUENCE OF 1421-1540 FROM N.A.  
 RX MEDLINE=94020842; PubMed=8414518;  
 RA Forster A., Rabbitts T.H.;  
 RT "A method for identifying genes within yeast artificial chromosomes:  
 application to isolation of MLL fusion cDNAs from acute leukaemia  
 translocations.";  
 RL Oncogene 8:3157-3160(1993).  
 RN [9]  
 RP CHROMOSOMAL TRANSLLOCATION WITH GAS7.  
 RX MEDLINE=20183971; PubMed=10706619;  
 RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,  
 RA Willson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,  
 RA Williams T.M., Lange B.J., Felix C.A.;  
 RT "Detection of leukemia-associated MLL-GAS7 translocation early during  
 chemotherapy with DNA topoisomerase II inhibitors.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).  
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.  
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLLOCATIONS  
 CC T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLLT1/ENL;  
 CC T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLLT2/AF4; T(9;11)(P22;Q23)  
 CC THAT INVOLVES MLL AND MLLT3/AF9; T(6;11)(Q27;Q23) THAT INVOLVES  
 CC MLL AND MLLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND  
 CC MLLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLLT7/AFX1;  
 CC T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLLT10/AF10;  
 CC T(1;11)(Q21;Q23) THAT INVOLVES MLL AND AF10; T(11;19)(Q23;P13.3)  
 CC THAT INVOLVES MLL AND ELL; AND T(11;19)(Q23;P23) THAT INVOLVES MLL  
 CC AND GAS7.  
 CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION  
 CC FACTORS.  
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 PHD ZINC-FINGER DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; L04284; AAA58669.1; -;  
 DR EMBL; Z69744; CA93625.1; -;  
 DR EMBL; Z69745; CA93625.1; JOINED.  
 DR EMBL; Z69746; CA93625.1; JOINED.  
 DR EMBL; Z69747; CA93625.1; JOINED.  
 DR EMBL; Z69748; CA93625.1; JOINED.  
 DR EMBL; Z69749; CA93625.1; JOINED.  
 DR EMBL; Z69750; CA93625.1; JOINED.  
 DR EMBL; Z69751; CA93625.1; JOINED.  
 DR EMBL; Z69752; CA93625.1; JOINED.  
 DR EMBL; Z69753; CA93625.1; JOINED.  
 DR EMBL; Z69754; CA93625.1; JOINED.  
 DR EMBL; Z69755; CA93625.1; JOINED.  
 DR EMBL; Z69756; CA93625.1; JOINED.  
 DR EMBL; Z69757; CA93625.1; JOINED.  
 DR EMBL; Z69758; CA93625.1; JOINED.  
 DR EMBL; Z69759; CA93625.1; JOINED.  
 DR EMBL; Z69760; CA93625.1; JOINED.  
 DR EMBL; Z69761; CA93625.1; JOINED.  
 DR EMBL; Z69762; CA93625.1; JOINED.  
 DR EMBL; Z69763; CA93625.1; JOINED.  
 DR EMBL; Z69764; CA93625.1; JOINED.  
 DR EMBL; Z69765; CA93625.1; JOINED.  
 DR EMBL; Z69766; CA93625.1; JOINED.  
 DR EMBL; Z69767; CA93625.1; JOINED.  
 DR EMBL; Z69768; CA93625.1; JOINED.  
 DR EMBL; Z69769; CA93625.1; JOINED.  
 DR EMBL; Z69770; CA93625.1; JOINED.  
 DR EMBL; Z69772; CA93625.1; JOINED.  
 DR EMBL; Z69773; CA93625.1; JOINED.  
 DR EMBL; Z69774; CA93625.1; JOINED.  
 DR EMBL; Z69775; CA93625.1; JOINED.  
 DR EMBL; Z69776; CA93625.1; JOINED.  
 DR EMBL; Z69777; CA93625.1; JOINED.  
 DR EMBL; Z69778; CA93625.1; JOINED.  
 DR EMBL; Z69779; CA93625.1; JOINED.  
 DR EMBL; Z69780; CA93625.1; JOINED.  
 DR EMBL; Z69781; CA93625.1; JOINED.  
 DR EMBL; L01986; BAA02511.1; -;  
 DR EMBL; U04737; AAA18644.1; -;  
 DR EMBL; S78570; AAB34770.1; -;  
 DR EMBL; X83604; CA58584.1; -;  
 DR EMBL; S66432; AAB28545.1; -;  
 DR EMBL; AF231998; AAG26332.2; ALT\_TERM.  
 DR TRANSFAC; T02337; -;  
 DR MIM; 159555; -;  
 DR InterPro; IPR001214; -;  
 DR InterPro; IPR001487; -;  
 DR InterPro; IPR001965; -;  
 DR InterPro; IPR002857; -;  
 DR Pfam; PF00628; PHD; 3.  
 DR Pfam; PF00856; SET; 1.  
 DR Pfam; PF02008; ZF-CXXC; 1.  
 DR PROSITE; PS50014; BROMODOMAIN\_2; 1.  
 DR PROSITE; PS0280; SET; 1.  
 KW Proto-oncogene; Chromosomal translocation; DNA-binding; Bromodomain;  
 KW Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;  
 KW Alternative splicing.  
 FT DOMAIN 17 102 ALA/GLY/SER-RICH.  
 FT DNA\_BIND 169 180 A.T HOOK (BY SIMILARITY).  
 FT DNA\_BIND 217 227 A.T HOOK (BY SIMILARITY).  
 FT DNA\_BIND 301 309 A.T HOOK (BY SIMILARITY).  
 FT ZN\_FING 1431 1482 PHD-TYPE 1.  
 FT ZN\_FING 1484 1533 PHD-TYPE 2.  
 FT ZN\_FING 1566 1627 PHD-TYPE 3.  
 FT DOMAIN 1703 1748 BROMODOMAIN (DIVERGENT).  
 FT DOMAIN 137 143 SET.  
 FT DOMAIN 561 564 POLY-GLY.  
 FT DOMAIN 568 571 POLY-PRO.  
 FT SITE 1444 1445 BREAKPOINT FOR TRANSLLOCATION TO FORM MLL-  
 FT VARSPLIT 1407 1444 GAS7 ONCOGENE.  
 FT MISSING (IN ISOFORM 14P-18B).





```

132 InArgLeuLeuGlyPro.SerPheSerHisGlyProAlaAlaAlaAspG1 148
102 CATAATGGAACACAGCTCCCTGCGCCCTTTGAGGGTGGCTGTCAAA..... 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 ylleleargValGlnPro.....LysSerGluLeuLysProGlyA 162
145 .....CATGCCCTCGA... 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
162 spValLysProLeuSerLysAspAspLeuGlyLeuHisAlaTyrArgCys 178
157 ....GAGCTGGTGAAGCGGTAGTTGTACAGACTAACAACGCCCAATGCG 203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 GluAspCysGlyLysCys....LysCysLysGluCysThrTyrProArgP 194
204 GTCGATGTTATCGGACCATTCGCTGTC.....TGTTTACTGTGTA 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 O.....LeuProSerAspTrpIleCysAspLysGlnCysLeu.CysSer 208
245 GCGCTTGATGACTTCCTGCTGCTACTCCGCTTGGAGCCATTGAGTGTG 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209 AlaGlnAsn.....ValIleAspTyrGlyThrCysValCysCy 221
295 TTTA..... 298
221 sValLysGlyLeuPheTyrHisCysSerAsnAspAspGluAspAsnCysA 238
144 .....CCGTGCAGCACTCAAGACTTCCTGCTGCTACTGCTGGCAG 340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 laAspAsnProCysSerCysSerGlnSerHisCysCysThrargTrpSer 254
341 .....CATPAGTCTGTTCTGCTGCTGCTGCTGGGACACCTGTGT 378
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 AlaMetGlyValMetSerLeuPheLeu.....ProCysLe 266
379 ACTACAGTACTTC.....GTTCCCTTTGTAGTGTGT 409
: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 utrpCysTyrLeuProAlaLysGlyCysLeuLysLeuCysGlnGly 281
seq_name: SwissProt_39:MR15_DROME
seq_documentation_block:
ID MR15_DROME STANDARD; PRT; 424 AA.
AC Q9Y0I1; Q9VF99;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MRG15 PROTEIN.
GN MRG15 OR CG6363.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Bertram M.J., Pereira-Smith O.M.;
RT "Conservation of the MORF4 related gene family: identification of a
RT new chromo domain subfamily and novel protein motif";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman D.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

```

```

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE MRG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF152245; AAD38047.1; -.
DR EMBL; AE003708; AAF55161.1; -.
DR FlyBase; FBgn0027378; MRG15.
KW Nuclear protein.
SQ SEQUENCE 424 AA; 47194 MW; B0EIF615252D8EDD CRC64;

alignment_scores:
Quality: 70.00 Length: 67
Ratio: 1.842 Gaps: 1
Percent Similarity: 56.716 Percent Identity: 34.328

alignment_block:
US-09-540-234-1/rev x MR15_DROME ..

Align seg 1/1 to: MR15_DROME from: 1 to: 424
417 GAAGTAGTACCACTACAAAGGAAACGAAGTACTGTAGTACACAGGTGTC 368
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 GluTrpValProGluAsnArgValLeuLysTyrAsnAspAspVal 82
367 CCCAGACGAGCGCAACACAGACTATGTCGCCGCGAGTA..... 328
: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
82 sargargGlnGluLeuAlaArgGlnCysGlyGluArgSerLysLysAspA 99
327 ..CACCAGGAGTCTTCGAGTGTGCAGCGTAAACAGCACTCAATGCG 280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
99 snLysLysGlySerAlaLysAlaLysLysMetGlnMetArgAsnGlu 115
279 TCACAGCGGATACAGCAGGAGTATCATCAAGCGCTACACAGTAACACAGC 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 SerArgAlaSerThrProSerLysAspSerAsnThrSerGlnSerThrAl 132
229 A 229

```

132 a 132

seq\_name: SwissProt\_39:GCR2\_YEAST

seq\_documentation\_block:

ID GCR2\_YEAST STANDARD; PRT; 534 AA.  
 AC Q01722;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE GLYCOPOLYIC GENES TRANSCRIPTIONAL ACTIVATOR GCR2.  
 GN GCR2 OR YNL199C OR N1374.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92375051; PubMed=1508187;  
 RA Uemura H., Jigami Y.;  
 RT "Role of GCR2 in transcriptional activation of yeast glycolytic  
 genes.";  
 RL Mol. Cell. Biol. 12:3834-3842(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=95242839; PubMed=7725799;  
 RA Jonniaux J.-L., Coster F., Fumelle B., Goffeau A.;  
 RT "A 21.7 kb DNA segment on the left arm of yeast chromosome XIV  
 carries WHI3, GCR2, SPX18, SPX19, an homologue to the heat shock gene  
 SSB1 and 8 new open reading frames of unknown function.";  
 RL Yeast 10:1639-1645(1994).  
 CC -1- FUNCTION: REQUIRED FOR THE EXPRESSION OF MOST GLYCOPOLYIC GENES.  
 CC GCR1 AND GCR2 PROBABLY FUNCTION AS A TRANSCRIPTIONAL ACTIVATION  
 CC COMPLEX. GCR1 PROVIDING THE SPECIFIC DNA-BINDING FUNCTION AND  
 CC GCR2 PROVIDING THE ACTIVATION FUNCTION.  
 CC -1- SUBUNIT: FORMS A GCR1/GCR2 COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- SIMILARITY: SOME, TO GCR1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

EMBL; D10104; BAA00985.1; -;  
 DR EMBL; X78898; CAA5509.1; -;  
 DR EMBL; Z71475; CAA96097.1; -;  
 DR PIR; S31300; S31300.  
 DR SGD; S0005143; GCR2.  
 KW Transcription regulation; Activator; Nuclear protein.  
 FT DOMAIN 281 288 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 255 283 ASN-RICH.  
 SQ SEQUENCE 534 AA; 58062 MW; 958D4A9393255B1B CRC64;

alignment\_scores:  
 Quality: 69.50 Length: 126  
 Ratio: 0.993 Gaps: 6  
 Percent Similarity: 55.556 Percent Identity: 23.810

alignment\_block:

US-09-540-234-1/rev x GCR2\_YEAST ..

Align seg 1/1 to: GCR2\_YEAST from: 1 to: 534

433 ACAGAGTACATCGCTAGCAAGTAGTACCACTACAAGGGAACGAGTACT 384  
 |||||  
 75 ThrSerThrGlySerAsnAlaLeuThrLysThrAsnSerAla.. 90

383 GTAGTACACAGGTGTCCCGACAGCAGGCAAGCAACAGACTATGCTG..... 338  
 91 .....AsnGlyThrProAsnGlyAsnSerSerSerThrSerAlaIles 105  
 337 .....CCAGCGAGTACAGCAAGAAAGTCTTCGAGTGTGCA 302  
 105 erAsnAlaSerAsnProAlaThrThrGlyAsnAsnAlaSerSerSerAla 121  
 301 CGGTAAACAGCACTCAATGGCTCCAGCGGAGTACAGCAAG..... 260  
 122 ThrSerAsnGlyIle...TyrThrGlnAlaGlnTyrSerGlnLeuPheAl 137  
 259 .GAAGTCATCAAGCGCTAC.....ACAGTAAACAGCACACGAATGGCTC 217  
 137 aLysIleSerLysLeuTyrAsnAlaThrLeuSerSerGlySerIleAspA 154  
 216 CGATAACATCGACGCCATTCGGCTGTGTAGTCTTGTACAAA..... 176  
 154 sPArgSerThrSerProLysSerAlaIleGluLeuTyrGlnArgPheGln 170  
 175 .....CTACG 171  
 171 GlnMetIleLysGluLeuGluLeuSerPheAspAlaSerProTyrAlaLy 187  
 170 GCTTCACCGACGTCTCGAGGGGCATGTT 143  
 187 sTyrPheArgArgLeuAspGlyArgLeu 196

seq\_name: SwissProt\_39:FIMC\_BORPE

seq\_documentation\_block:

ID FIMC\_BORPE STANDARD; PRT; 873 AA.  
 AC P33410;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE OUTER MEMBRANE USHER PROTEIN FIMC PRECURSOR.  
 GN FIMC OR FHAA.  
 OS Bordetella pertussis.  
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 OC Bordetella.  
 OX NCBI\_TaxID=520;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=WELLCOME 28;  
 RX MEDLINE=93078620; PubMed=1360139;  
 RA Williams R.J.L., van der Heide H.G.J., Mooi F.R.;  
 RT "Characterization of a Bordetella pertussis fimbrial gene cluster  
 RT which is located directly downstream of the filamentous  
 RT haemagglutinin gene.";  
 RL Mol. Microbiol. 6:2661-2671(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TOHAWA I;  
 RX MEDLINE=92371423; PubMed=1354611;  
 RA Loch C., Geoffroy M.C., Renaud G.;  
 RT "Common accessory genes for the Bordetella pertussis filamentous  
 RT hemagglutinin and fimbriae share sequence similarities with the papC  
 RT and papD gene families.";  
 RL EMBO J. 11:3175-3183(1992).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

```
DR EMBL; X64876; CAA46090.1; -.
DR EMBL; X66729; CAA47266.1; -.
DR PIR; S21574; S21574.
DR PIR; S25193; S25193.
DR InterPro; IPR000015; -.
DR Pfam; PF00577; Usher; 1.
KW POSITIVE; PS01151; FIMBRIAL_USHER; 1.
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 873 OUTER MEMBRANE USHER PROTEIN FIMC.
FT DISULFID 815 938 POTENTIAL.
FT CONFLICT 744 744 G -> A (IN REF. 2).
SQ SEQUENCE 873 AA; 93369 MW; AC88D8078388CB22 CRC64;

alignment_scores:
  Quality: 69.00 Length: 123
  Ratio: 0.986 Gaps: 4
  Percent Similarity: 56.911 Percent Identity: 25.203

alignment_block:
  US-09-540-234-1/rev x FIMC_BORPE ..

Align seg 1/1 to: FIMC_BORPE from: 1 to: 873
408 CCACCTACAAAGGAAAGTACTGTAGTACACAGGTGTCCCGACCA 359
||||:||||:||||:||||:||||:||||:||||:||||:||||:
597 ProValGlnSerValAsnTyrSerGlyMetSerGlyGluArgAspG1 613
358 GGCAGCAACAGACTATGTCGCCAGGACGATACAGCAAGGAGTCTTCGAG 309
|:||||:||||:||||:||||:||||:||||:||||:||||:
613 nTyrThrTyrGlyLeuGlyValGlnArgAlaGlyThrSerAlaGlnTyrA 630
308 TGCTGCAGCGTAAACAGCACTCAATGGCTCCCAAGCGGAGTACAGCAAGG 259
||||:||||:||||:||||:||||:||||:||||:||||:||||:
630 lAluasnGlySerTrpSerGlyThrTyrGlyGluValSer.GlyGlnLe 646
258 AAGTCATCAAGCGCTACACAGTAGTAACAGCAGCAGATGGCTCCGATAACA 209
||||:||||:||||:||||:||||:||||:||||:||||:||||:
646 uThrHisGlyArgSerTyrSerGlnTyrGlnIleAsnGlySerGlyGlyL 663
208 TCAGACGCC.....ATGGCGGTT 192
663 euValAlaHisAlaGlyValThrPheGlyGlnTyrGlnAlaGlyThr 679
191 GTTAGTCTTCAAACTACGGCTTCACCG.....ACGTC 157
680 lIleGlyLeuIleGlnAlaGluAlaAlaGlyAlaLysValValAsnTh 696
156 TCGAGGGCGCATGTTTGAGCAGCACCCCTCAAAAGGCGAGGACTGTTTCCA 107
||||:||||:||||:||||:||||:||||:||||:||||:||||:
696 rArgAsnAlaAlaValAspArg.....SerGlyTyrGlyLeuValSerL 711
106 TTAGTGGCGTCCAGCTA 90
:: ||| ||| ||| |||
711 euthrProTyrSerLeu 716

seq_name: SwissProt_39:M84D_DROME

seq_documentation_block:
ID M84D_DROME STANDARD; PRT; 68 AA.
AC Q01645;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MALE SPECIFIC SPERM PROTEIN MST84DD.
GN MST84DD.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]

RP SEQUENCE FROM N.A.
RX STRAIN=OREGON-R; PubMed=1684716;
RA Kuhn R., Kuhn C., Boersch D., Glaetzer K.H., Schaefer U.,
RA Schaefer M.;
RT "A cluster of four genes selectively expressed in the male germ line
of Drosophila melanogaster.";
RL Mech. Dev. 35:143-151(1991).
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -!- DEVELOPMENTAL STAGE: PRIMARY SPERMATOCYTES.
CC -!- DOMAIN: THIS PROTEIN IS MOSTLY COMPOSED OF REPETITIVE C-G-P
MOTIFS.
CC -!- SIMILARITY: BELONGS TO THE MST(3)CGP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; X67703; CAA47940.1; -.
DR HSSP; P01180; INPO.
DR FlyBase; FBgn0004175; Mst84Dd.
KW Spermatogenesis; Repeat; Multigene family.
SQ SEQUENCE 68 AA; 6480 MW; 2F2BD26128DE3DEF CRC64;

alignment_scores:
  Quality: 68.50 Length: 33
  Ratio: 3.806 Gaps: 3
  Percent Similarity: 54.545 Percent Identity: 48.485

alignment_block:
  US-09-540-234-1 x M84D_DROME ..

Align seg 1/1 to: M84D_DROME from: 1 to: 68
229 TGCTGTTTACTGTAGCGCTTGATGACITCCCTGCTGCTACTCGCTTGG 278
||||| ||| ||| ||| ||| ||| ||| ||| |||
20 CysCysGlyProCysGly.....ProCysCysGlyProCysG1 33
279 AGCAATTGAGTGCTGTTTACCGTCGACGACTCGAAGACTTCCTTCG 325
| ||| ||||| ||||| ||| ||| ||| |||
33 y.Pro.....CysCysGlyProCysGlyProArgCysGlyProCys 46

seq_name: SwissProt_39:TBAL_PELFA

seq_documentation_block:
ID TBAL_PELFA STANDARD; PRT; 453 AA.
AC Q40831;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TUBULIN ALPHA-1 CHAIN.
GN TUBAL.
OS Pelvetia fastigiata.
OC Eukaryota; stramenopiles; Phaeophyceae; Fucales; Pelvetia.
OX NCBI_TaxID=48072;
RN [1]
RP SEQUENCE FROM N.A.
RA Cofman H.R., Kropf D.L.;
RT "The brown alga, Pelvetia fastigiata, expresses two alpha-tubulin
sequences.";
RL (in) Plant Gene Register PCR97-019.
CC -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC -!- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.
CC -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
```





```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=6957;
RX  MEDLINE=90291515; PubMed=2357738;
RA  Shumard-Hudspeth D.S., Hudspeth M.E.;
RT  "Genic rearrangements in Phytophthora mitochondrial DNA.";
RL  Curr. Genet. 17:413-415(1990).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=6957;
RA  Sachay D.J., Hudspeth D.S., Nadler S.A., Hudspeth M.E.;
RT  "Oomycete mtDNA: Phytophthora genes for cytochrome c oxidase use an
RT  unmodified genetic code and encode proteins most similar to plants.";
RL  Mycol. 17:7-23(1993).
CC  -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC  CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC  3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC  CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC  CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC  AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3
CC  AND COPPER B.
CC  -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC  4 FERRICYTOCHROME C.
CC  -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC  -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC  INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC  -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; L04457; AAA32023.2; -
DR  HSP; P00396; LOCC.
DR  InterPro; IPR000883; -
DR  InterPro; IPR002428; -
DR  Pfam; PF00115; COX1; 1.
DR  PRINTS; PR01165; CYC1DASE1.
DR  PROSITE; PS00077; COX1; 1.
KW  Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW  Respiratory chain; Inner membrane.
FT  METAL 70 70 IRON (HEME A) (PROBABLE).
FT  METAL 249 249 COPPER B (PROBABLE).
FT  METAL 253 253 COPPER B (PROBABLE).
FT  METAL 298 298 COPPER B (PROBABLE).
FT  METAL 299 299 COPPER B (PROBABLE).
FT  METAL 384 384 IRON (HEME A3) (PROBABLE).
FT  METAL 386 386 IRON (HEME A3) (PROBABLE).
SQ  SEQUENCE 492 AA; 54325 MW; 9214D85A2E32021 CRC64;

```

```

alignment_scores:
  Quality: 67.00      Length: 149
  Ratio: 0.971      Gaps: 8
Percent Similarity: 46.309      Percent Identity: 27.517

```

```
alignment_block:
US-09-540-234-1 x COX1_PHYME ..

```

```
Align seg 1/1 to: COX1_PHYME from: 1 to: 492

```

```

107 TGGAAACACTCCCTGCTTTAGGGTCTCAAAATGCCCTCGA 156
    ||| ||||| ||||| ||||| ||||| ||||| |||||
112 TrpLeuLeuProAlaLeuLeuValSerAlaLeuValG 128
157 GAGC.....TCGGTGAAGCCGTAGTTT.....GTAC 182
    ||||| ||||| ||||| ||||| ||||| |||||
128 userGlyAlaGlyThrGlyTrpThrValTyProProLeuSerValG 145

```

```

183 AAGACTTAACAGCCCAATGGCGTCGAT..... 209
    ||| ||||| ||||| ||||| ||||| |||||
145 In.AlahisSerGlyProSerValasPheLeuAlaIlePheSerLeuHisLe 161
210 .....GTTATCGAGCCATTTCGTGCTGCTTTACTGTGT 243
    ||||| ||||| ||||| ||||| ||||| |||||
161 uThrGlyIleSerSerLeuLeuGlyAlaIleAsnPheIleSerThrIle 178
244 AGCGCTTGATGACT.....TCCTTGCTGTACTCCGCT 275
    ||||| ||||| ||||| ||||| ||||| |||||
178 yrAsnMetArgAlaProGlyLeuSerPheHisArgLeuProLeuPheVal 194
276 TGGAGCATTGAGTGTGCTTTACCGTCGACACTCGAAGACTTCCTTGC 325
    ||||| ||||| ||||| ||||| ||||| |||||
195 TrpSerValLeuIleThrAlaPheLeuLeuLeuThr.....LeuPr 209
326 TGCTACTCGCTGGCAGCATAGTCGTTTGGCTG..... 356
    ||||| ||||| ||||| ||||| ||||| |||||
209 oValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnLeuAsnT 226
357 .....CCGTGCTGGGACACCTGTGTACTCAGTACT 389
    ||||| ||||| ||||| ||||| ||||| |||||
226 hrSerPheTyraPpProSerGlyGlyAspProValLeuTyGlnHis 242
390 TCCTTT.....CCCTTGTAGTGTACTACTTCTT 419
    ||||| ||||| ||||| ||||| ||||| |||||
243 LeuPheTrpPheGlyHisProGluValTyIleLeuLeuLeu 257

```

```
seq_name: SwissProt_39:COX1_LOCMI
```

```
seq_documentation_block:
```

```

ID COX1_LOCMI STANDARD; PRT; 512 AA.
AC Q36421;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
OS COI.
OS Locusta migratoria (Migratory locust).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC Acridomorpha; Acridoidea; Acrididae; Locusta.
OX NCBI_TaxID=7004;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96139026; PubMed=8587138;
RA Flook P.K., Rowell C.H.F., Gellissen G.;
RT "The sequence, organization, and evolution of the Locusta migratoria
RT mitochondrial genome.";
RL J. Mol. Evol. 41:928-941(1995).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```







```
CC INHIBITION OR DAMAGE BY O(2) UNDER CARBON- OR PHOSPHATE-LIMITED
CC CONDITIONS.
CC -1- CATALYTIC ACTIVITY: 2 REDUCED FERREDOXIN + 2 H(+) -> 2 OXIDIZED
CC FERREDOXIN + H(2).
CC -1- COFACTOR: CONTAINS NICKEL AND IRON-SULFUR.
CC -1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -1- SIMILARITY: BELONGS TO THE [NIFE]/[NIFESE] HYDROGENASE LARGE
CC SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S56898; JAB25780.1; ALT_SEQ.
DR PIR; JH0776; JH0776.
DR HSSP; P12944; IFRV.
DR InterPro; IPR001501; -.
DR Pfam; PF00374; Nifese_Hases; 1.
DR PROSITE; PS00507; NI_HGENASE_L1; FALSE_NEG.
DR PROSITE; PS00508; NI_HGENASE_L2; 1.
KW Oxidoreductase; Membrane; Nickel; plasmid.
FT METAL 75 75 NICKEL (POTENTIAL).
FT METAL 78 78 NICKEL (POTENTIAL).
FT METAL 600 600 NICKEL (POTENTIAL).
FT METAL 603 603 NICKEL (POTENTIAL).
SQ SEQUENCE 621 AA; 68924 MW; 906AC4414285D4A3 CRC64;

alignment_scores:
  Quality: 67.00 Length: 88
  Ratio: 1.634 Gaps: 6
  Percent Similarity: 46.591 Percent Identity: 31.818

alignment_block:
US-09-540-234-1 x MBHL_ALCHY ..

Align seg 1/1 to: MBHL_ALCHY from: 1 to: 621

37 TGGGTCACTGAAGGAAGGCAAGACACACCTCCAAAGCGGACGGCTA 86
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
519 TrpProLysGluAlaLysGlyValGlyHisValAlaAlaProArgGlyAla 535
87 CRTGAGC...TGG.....CACGGCATAATGGAACACAGTCCCT 121
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
535 acysGlyHisTrpIleArgIleLysAspGlyLysIleGluAsnTyrGlnC 552
122 GCCCTTTGAGGGTGTCTCAACATGCCCCCTCGAGAC..... 159
|| || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
552 ysValValProThrThrTrpAsnGlySerProArgAspSerLysGlyGln 568
160 GTCGGTGAAGCGGTAGTTGTGTACAGACTAACACGCCAATCGCGTCGAT 209
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
569 IleGly...AlaPheGluAlaSerLeuMetAsnThrProMetAlaLysPr 584
210 GTTATCGGAGCCA.....TTTCGTGTGCT 232
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
584 o...GluGluProValGluIleLeuArgThrValHisSerPheAspProc 600

233 GTTACTGTGTAGC 246
||||| |||||
600 ysLeuAlaCysSer 604

seq_name: SwissProt_39:FABB_HAEIN

seq_documentation_block:
ID FABB_HAEIN STANDARD; PRT; 406 AA.
AC P43710;
DT 01-NOV-1995 (Rel. 32, Created)
```

```
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 3-OXOACYL-[ACYL-CARRIER-PROTEIN] SYNTHASE I (EC 2.3.1.41) (BETA-
DE KETOACYL-ACP SYNTHASE I) (KAS I).
DE FABB OR H11533.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uutterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: CATALYZES THE CONDENSATION REACTION OF FATTY ACID
CC SYNTHESIS BY THE ADDITION TO AN ACYL ACCEPTOR OF TWO CARBONS
CC FROM MALONYL-ACP. SPECIFIC FOR ELONGATION FROM C-10 TO UNSATURATED
CC C-16 AND C-18 FATTY ACIDS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ACYL-[ACYL-CARRIER PROTEIN] + MALONYL-[ACYL-
CC CARRIER PROTEIN] = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + CO(2) +
CC [ACYL-CARRIER PROTEIN].
CC -1- PATHWAY: FATTY ACID BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE BETA-KETOACYL-ACP SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32829; AAC23183.1; -.
DR TIGR; H11533; -.
DR InterPro; IPR000794; -.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
KW Fatty acid biosynthesis; Transferase; Acyltransferase.
FT ACY_SITE 162 162 BY SIMILARITY.
SQ SEQUENCE 406 AA; 42649 MW; A17B8673156FDDAF CRC64;

alignment_scores:
  Quality: 66.50 Length: 118
  Ratio: 1.090 Gaps: 5
  Percent Similarity: 51.695 Percent Identity: 25.424

alignment_block:
US-09-540-234-1/rev x FABB_HAEIN ..

Align seg 1/1 to: FABB_HAEIN from: 1 to: 406

440 AACAGACACGAGTACATGCTAAGAGTAGTACCCTACCAAGGAAACG 391
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 AsnProSerGluHisIleAspArgLysValPheArgPheMetGly.... 69
390 AAGTACTGTAGTACACAGCTGTCCCGACGACCCAGCCAGCAACAGACTATG 341
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
70 .....AspAla 72
```

340 CTGCCAGCGAGTAC.....AGCAAGGAAGTCTTCGAGTGTCTGCACGGTA 297  
 ||||| ||||| :|||:|||||:||||| :||| :||| :|||  
 72 laAlaTyraLalTyLeuSerMetArgluAlaIleGluAspAlaGlyLeu 88  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 296 AACGACACTCAATGGCTCAAGC.....GGAGTACA 265  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 89 ThrGluAspGlnValSerAsnAspArgThrGlyLeuValIleGlyAlaG1 105  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 264 GCAAGGAAGT...CATCAAGCGGTACACAGTAACACACACACGAATGGCT 218  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 105 yThrGlySerAlaHisnGlnLeuValAlaCysAspAlaValArgGlyP 122  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 217 CCGATACATCGACGCGCATGGCGTGTAGTCTTGTACAACTACGCGCT 168  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 122 roArgGlyValIleAlaIleGlyProTyrAlaValThrGlyThrMetAla 138  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 167 TCACCGACGCTCGAGGGCGATGTTGACAGCAGCCCTCAAAGGGCAGG 118  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 139 SerSerValSer.....AlaCysLeuAlaThrProTyrLysIleArgG1 153  
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
 117 ACTG 114  
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
 153 yVal 154

seq\_name: SwissProt\_39:FRA\_DROME

seq\_documentation\_block:  
 ID FRA\_DROME STANDARD; PRT; 595 AA.  
 AC P21525;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE TRANSCRIPTION FACTOR DFRA (FOS-RELATED ANTIGEN) (AP-1) (KAYAK  
 DE PROTEIN).  
 GN KAY OR FRA  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=90337318; PubMed=2116361;  
 RA Perkins K.K., Admon A., Patel N., Tjian R.;  
 RT "The Drosophila Fos-related AP-1 protein is a developmentally  
 RT regulated transcription factor.";  
 RL Genes Dev. 4:822-834(1990).  
 CC -!- FUNCTION: DEVELOPMENTALLY REGULATED TRANSCRIPTION FACTOR AP-1  
 CC BINDS AND RECOGNIZES THE ENHANCER DNA SEQUENCE: TGA(C/G)TCA. MAY  
 CC PLAY A ROLE IN THE FUNCTION OR DETERMINATION OF A PARTICULAR  
 CC SUBSET OF CELLS IN THE DEVELOPING EMBRYO. IT IS ABLE TO CARRY OUT  
 CC ITS FUNCTION EITHER INDEPENDENTLY OF OR IN CONJUNCTION WITH DJRA.  
 CC -!- SUBUNIT: MAY FORM DIMER OF IDENTICAL CHAINS AND MAY ALSO INTERACTS  
 CC WITH DJRA TO FORM AN HETERODIMER. DJRA-DJRA COMPLEX IS BOUND MORE  
 CC STABLY TO THE AP-1 SITE THAN EITHER OF THE TWO PROTEINS ALONE.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- TISSUE SPECIFICITY: CELLS TYPE OF THE EMBRYO THAT ARE INVOLVED IN  
 CC THE DEVELOPMENT OF THE HEAD AND NERVOUS SYSTEM.  
 CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. FOS SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; X54143; CAA38082.1; -;  
 CC PIR; A35847; A35847.  
 CC HSSP; P01100; IFOS.  
 CC TRANSFAC; T01997; -;  
 CC FlyBase; FBgn0001297; kay.

DR InterPro; IPR001871; -;  
 DR Pfam; PF00170; bzip; 1.  
 DR PROSITE; PS00036; BZIP\_BASIC; 1.  
 KW Transcription regulation; DNA-binding; Activator; Nuclear protein.  
 FT DNA\_BIND 260 314 BASIC MOTIF.  
 FT DOMAIN 286 314 LEUCINE-ZIPPER.  
 SQ SEQUENCE 595 AA; 62837 MW; C4D81F99582D38DE CRC64;

alignment\_scores:  
 Quality: 66.50 Length: 69  
 Ratio: 1.357 Gaps: 2  
 Percent similarity: 71.014 Percent Identity: 26.087

alignment\_block:

US-09-540-234-1/rev x FRA\_DROME ..

Align seg 1/1 to: FRA\_DROME from: 1 to: 595

357 GCAGCAACAGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGT 308  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 319 AlaThrHisArgAlaThrCysGlnLysIleArgSerAspMetLeuSerVa 335  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 307 GCTCCACGGTAAACAGCAGCTCAATGGCTCCACGGGAGTACACAGGA 258  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 335 lVal.ThrCysAsnGlyLeu...IleAlaProAlaGlyLeuLeuSerAla 350  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 257 AGTCATCAAGCGGTACACAGTAAACAGCACACAGTAATGGCTCCGATAACAT 208  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 351 GlySerSerGlySerGlyAlaSerSerHisHisnHisnSerAsnAs 367  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 207 CGAGCGCATGGCGTGTAGTCTTGTACAACTACGCGCTTCACGCGT 158  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 367 pSerSerAsnGlyThrIleThrGlyMetAspAlaThrLeuAsnSerThrG 384  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 157 CTCGA 153  
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
 384 lyArg 385

seq\_name: SwissProt\_39:PRIA\_BACSU

seq\_documentation\_block:

ID PRIA\_BACSU STANDARD; PRT; 805 AA.  
 AC P94461; O34941;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y).  
 GN PRIA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA "DNA sequence of a 28 Kbp segment of DNA from the spvM region of  
 RT Bacillus subtilis.";  
 RT Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RN SEQUENCE OF 230-895 FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97440812; PubMed=9086272;  
 RA Mazel D., Coic E., Blanchard S., Saurin W., Marliere P.;  
 RT "A survey of polypeptide deformylase function throughout the  
 RT eubacterial lineage.";  
 RL J. Mol. Biol. 266:939-949(1997).  
 CC -!- FUNCTION: RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX SSDNA;  
 CC THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS PRIB AND  
 CC PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE SUBSEQUENT  
 CC ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN FUNCTIONS AS A  
 CC HELICASE WITHIN THE PRIMOSOME (BY SIMILARITY).  
 CC





RP SEQUENCE FROM N.A.  
 RA Zardoya R.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B.  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +  
 CC 4 FERRICYTOCHROME C.  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL  
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF039066; AAD05052.1; -.  
 DR InterPro; IPR000883; -.  
 DR InterPro; IPR002428; -.  
 DR Pfam; PF00115; COX1; 1.  
 DR PRINTS; PR01165; CYCOXIDASE1.  
 DR PROSITE; PS00077; COX1; 1.  
 DR OXIDOREDUCTASE; Heme; Copper; Mitochondrion; Transmembrane;  
 KW Respiratory chain; Inner membrane.  
 FT METAL 61 61 IRON (HEME A) (PROBABLE).  
 FT METAL 240 240 COPPER B (PROBABLE).  
 FT METAL 244 244 COPPER B (PROBABLE).  
 FT METAL 290 290 COPPER B (PROBABLE).  
 FT METAL 291 291 COPPER B (PROBABLE).  
 FT METAL 376 376 IRON (HEME A3) (PROBABLE).  
 FT METAL 378 378 IRON (HEME A) (PROBABLE).  
 SQ SEQUENCE 514 AA; 56977 MW; 195105D81C814E81 CRC64;

alignment\_scores:  
 Quality: 66.00 Length: 144  
 Ratio: 0.892 Gaps: 9  
 Percent Similarity: 51.389 Percent Identity: 27.083

alignment\_block:  
 US-09-540-234-1 x COX1\_PELSU ..  
 Align seg 1/1 to: COX1\_PELSU from: 1 to: 514

92 GCTGGCAGCGCATATGGAACAGTCCTCCCTTTTGGGGTCTGCTC 141  
 122 AlaglyThrGly...TrpThrValtyrProLeu.....Alase 134  
 142 AAACATGCC...CTCGACGCTCGGTGAAGCGGTAGTTGTACAGACT 188  
 134 rAsnLeuAlaHisAlaglyAlaserVal.AspleuAlaIlePheSerLeu 150  
 189 AACACGCCAATGGCTGATGTTATCGGAGCCATTGCTGCTGTTTAC 238  
 151 HisLeuAlaGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleThr 167  
 239 TGTGTAGCCCTGTGAGTACT.....TCCTGCTGTGACT 270  
 167 rValIleAsnMetLysThrProAsnMetSerPheLeuAspMetProLeu 184  
 271 CGCCTGGAGCATTTGAGTGTGCTTTTACCGTCGACCACTCGAAGACT 320  
 184 heValtrpSerValLeuIleThrAlaIleLeuLeuLeuSer..... 198

321 CTTGCTGTACTCGTCGGCAGCATAGTCTCTTGTCTG..... 356  
 199 LeuProValLeuAlaAlaglyIleThrMetLeuLeuThrAspArgAsnLe 215  
 357 .....CTGGTCTGGGGACACCTGTGTACTACA 384  
 215 uAsnThrThrPhePheAspProSerGlyGlyGlyAspProIleLeuTyrg 232  
 385 GTACTTCGTTT.....CCCTTTGTAGTGTGCTACTACTTCTT 419  
 232 InHisLeuPheTrpPhePheGlyHisProGluValtyrIleLeuIleLeu 248  
 420 .....AGCCATGTACTC 431  
 249 ProGlyPheGlyIleIleSerHisIleVal 258  
 seq\_name: SwissProt\_39:COX1\_MARPO  
 seq\_documentation\_block:  
 ID COX1\_MARPO STANDARD; PRT; 522 AA.  
 AC P26856;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 GN COX1 OR COXI.  
 OS Marchantia polymorpha (Liverwort).  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Embryophyta; Marchantiophyta; Marchantiales;  
 OC Marchantiaceae; Marchantia.  
 OX NCBI\_TaxID=3197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=92114051; PubMed=1731062;  
 RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Norato N.,  
 RA Akashi K., Kanegae T., Ogura Y., Kohchi T., Ohyama K.;  
 RT "Gene organization deduced from the complete sequence of liverwort  
 RT Marchantia polymorpha mitochondrial DNA. A primitive form of plant  
 RT mitochondrial genome."  
 RL J. Mol. Biol. 223:1-7(1992).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B.  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +  
 CC 4 FERRICYTOCHROME C.  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL  
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M68929; AAC09451.1; -.  
 DR PIR; S25956; S25956.  
 DR HSP; P00396; 10CC.  
 DR Mendel; 2055; MARPO;cox1;1.  
 DR InterPro; IPR000883; -.  
 DR InterPro; IPR002428; -.  
 DR Pfam; PF00115; COX1; 1.  
 DR PRINTS; PR01165; CYCOXIDASE1.  
 DR PROSITE; PS00077; COX1; 1.  
 DR OXIDOREDUCTASE; Heme; Copper; Mitochondrion; Transmembrane;  
 KW Respiratory chain; Inner membrane.

FT	METAL	65	65	IRON (HEME A) (PROBABLE).
FT	METAL	244	244	COPPER B (PROBABLE).
FT	METAL	248	248	COPPER B (PROBABLE).
FT	METAL	293	293	COPPER B (PROBABLE).
FT	METAL	294	294	COPPER B (PROBABLE).
FT	METAL	379	379	IRON (HEME A3) (PROBABLE).
FT	METAL	381	381	IRON (HEME A) (PROBABLE).
SQ	SEQUENCE	522 AA;	57551 MW;	4B8BE1F03A11D01CRC64;

alignment_scores:		
Quality:	66.00	Length: 159
Ratio:	0.917	Gaps: 7
Percent Similarity:	45.283	Percent Identity: 22.642

alignment block:

US-09-540-234-1 x COX1 MARPO

Align seq 1/1 to: COX1\_MARPO from: 1 to: 522

107 TGGAAACAGTCCTCCCTTTTGAGGGTCGTCTCAACATGCCCTCGA 156  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 107 TrpLeuLeuProSerLeuLeuLeuLeuLeuLeuLeuValG1 123  
 157 GACG..... 160  
 |||||  
 123 uValG1cysGlySerGlyTrpThrValTyProProLeuSerGlyLeu 140  
 161 .....TCGGTGAGCCGTAGTTGTACAGACTAACA 193  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 140 hrSerHisSerGlyGlySerVal AspLeuAlaIlePheSerLeuHisLe 156  
 194 CGCCAAATGGGCTCGATGTTATCGGAGGCATCTCGTGTCTTACHTGT 243  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 156 uSerGlyValSerSerIleLeuGlySerIleAsnPheIleThrThrIleP 173  
 244 AGCGCTTG.....ATGACTTCCTTGCTGTACTCCGCT 275  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 173 heAsnMetArgAlaProGlyLeuThrMetHisArgLeuProLeuPheVal 189  
 276 TGGAGCCATTGTGAGTGTGTTACCGTGCAGACTCGAAGACTTCCTTGC 325  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 190 TrpSerValLeuValThrAlaPheLeuLeuLeuLeuSer.....LeuPr 204

326 TGTACTCGCTGGCAGCATAGTCTGTTTGCT ..... 356  
||||| ||||| ||||| :||| :||| :||| :||| :|||  
204 ovalleuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnT 221  
||||| ||||| ||||| :||| :||| :||| :||| :|||  
357 ..... CCTGGTCTGGGGACACTGTCAGTACTACT 389  
||| ||| ||| :||| :||| :||| :||| :|||  
221 hrThrPhePheAspProAlaGlyGlyAspProIleLeuTyGlnHis 237  
TCGTTT ..... CCCTTTGTAGTGGTACTACTCTT ..... 419  
||| ||| ||| :||| :||| :||| :||| :|||  
238 LeuPheTrpPheGluHisProGluValTyrlleuLeuLeuProG 254

seq\_name: **Swissprot\_39:COX1\_BETVU**

seq\_documentation\_block:

ID	COX1_BETVU	STANDARD;	PRT;	524 AA.
----	------------	-----------	------	---------

AC P24794;

DT 01-MAR-1992 (Rel. 21, Created)

DT	01-FEB-1996 (Rel. 33, Last sequence update)
DT	01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).

GN COXI OR COXI.

Beta vulgaris  
Mitochondrion

OC Mitochondrion.  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;  
Caryophyllales; Chenopodiaceae; Beta.  
NCBI\_TaxID=3555;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=CV. TK81-0;  
RC MEDLINE=9130331; PubMed=1651175;  
RA Senda M., Harada T., Mikami T., Sugura M., Kinoshita T.;  
RX "Genomic organization and sequence analysis of the cytochrome oxidase  
RT subunit II gene from normal and male-sterile mitochondria in sugar  
RT beet.";  
RL Curr. Genet. 19:175-181(1991).  
RN [2]  
RN SEQUENCE FROM N.A.  
RA Harada T., Mikami T., Kinoshita T.;  
RP "Nucleotide sequence of cytochrome c oxidase subunit I gene of sugar  
RT beet mitochondria.";  
RL Proc. Sugar Beet Res. 29:15-21(1987).  
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
CC AND COPPER B.  
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +  
CC 4 FERRICYTOCHROME C.  
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
DR EMBL; X57693; CAA040874.1; -;  
DR EMBL; M57645; AAA87330.1; -;  
DR HSSP; P00396; IOCC.  
DR Mendel; 5481; BETvu;cox1.1.  
DR InterPro; IPR000883; -;  
DR InterPro; IPR002428; -;  
DR Pfam; PF00115; COX1; 1.  
DR PRINTS; PR01165; CYCXIDASE1.  
DR PROSITE; PS00077; COX1; 1.  
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
KW Respiratory chain; Inner membrane.  
DR FT METAL 64 64 IRON (HEME A) (PROBABLE).  
DR FT METAL 243 243 COPPER B (PROBABLE).  
DR FT METAL 247 247 COPPER B (PROBABLE).  
DR FT METAL 292 292 COPPER B (PROBABLE).  
DR FT METAL 293 293 COPPER B (PROBABLE).  
DR FT METAL 378 378 IRON (HEME A3) (PROBABLE).  
DR FT METAL 380 380 IRON (HEME A3) (PROBABLE).  
DR FT CONFLICT 2 9 TNLVRLWF -> AV (IN REF. 2).  
DR FT CONFLICT 220 226 TTFEDPA -> RFLIRW (IN REF. 2).  
DR FT CONFLICT 229 229 MISSING (IN REF. 2).  
DR FT CONFLICT 501 503 EWM -> ND (IN REF. 2).  
SQ SEQUENCE 524 AA; 57580 MW; A862089E5C476EA CRC64;

alignment\_scores:

Quality: 66.00

Ratio: 1.

Percent Similarity: 49

•  
•  
•  
•  
•  
•

```
alignment_block:
  --c 00 540 004 1
```

US-09-540-234-1 x COX1\_BETVU

Align seg 1/1 to: COX1\_BETVU from: 1 to: 524

```
198 AATGGCGTC...GATGTTATCGAGCCATTCGTGCTGTACTGTGTA 244
:::|||||:::|||||:::|||||:::  ::: |||:::
156 SerGlyValSerSerIleLeuGlySerIleAsnPheIleThrThrIlePh 172
:::|||||:::|||||:::|||||:::  ::: |||:::
245 GCGCTTG.....ATGACTTCCTGTGTACTCCGCTT 276
:::|||||:::|||||:::|||||:::  ::: |||:::
172 eAsnMetArgGlyProGlyMetThrMetHisArgLeuProLeuPheValt 189
:::|||||:::|||||:::|||||:::  ::: |||:::
277 GGAGCCATTGAGTCTGTTTACCGTGCAGCAGCTCGAGACTTCCTGCT 326
||||| ||| ::::::::::|:::|||||::: |||
189 rpSerValLeuValThrAlaPheLeuLeuLeuSer.....LeuPro 203
||||| ||| ::::::::::|:::|||||::: |||
327 GTACTCGCTGCGCAGCATAGTCTGTTTGTG..... 356
||||| ||| ::::::::::|:::|||||::: |||
204 ValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnTh 220
||||| ||| ::::::::::|:::|||||::: |||
357 .....CCTGGTCTGGGACACCTGTGTACTACACTT 390
||||| ||| ::::::::::|:::|||||::: |||
220 rThrPhePheAspProAlaGlyGlyAspProIleLeuTyrGlnHisL 237
||||| ||| ::::::::::|:::|||||::: |||
391 CGTTT.....CCCTTTGAGTGGTACTACTTCTT..... 419
||||| ||| ::::::::::|:::|||||::: |||
237 euPheTrpPheGlyHisProGluValTyrIleLeuLeuProGly 253
||||| ||| ::::::::::|:::|||||::: |||
420 .....AGCCATGTACTC 431
||||| ||| ::::::::::|:::|||||::: |||
254 PheGlyIleIleSerHisIleVal 261
||||| ||| ::::::::::|:::|||||::: |||
seq_name: SwissProt_39:COX1_ORYSA
```

seq\_documentation\_block:

```
ID COX1_ORYSA STANDARD; PRT; 524 AA.
AC P14578;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1 OR COXI.
OS Oryza sativa (Rice).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90016815; PubMed=2552410;
RA Kadowaki K., Suzuki T., Kazama S., Oh-Fuchi T., Sakamoto W.;
RT "Nucleotide sequence of the cytochrome oxidase subunit I gene from
RL rice mitochondria."
RL Nucleic Acids Res. 17:7519-7519(1989).
RN [2]
RP SEQUENCE OF 1-57 FROM N.A.
RX STRAIN=CV, TAICHUNG 65; TISSUE=Shoot;
RX MEDLINE=92035104; PubMed=1718614;
RA Suzuki T., Kazama S., Hirai A., Akihama T., Kadowaki K.;
RT "The rice mitochondrial nad3 gene has an extended reading frame at
RT its 5' end; nucleotide sequence analysis of rice trnS, nad3, and
RT rps12 genes."
RL Curr. Genet. 20:331-337(1991).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
```

```
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15990; CAA34122.1; -.
DR EMBL; M57903; AAA70312.1; -.
DR PIR; S06761; ODRZ1.
DR HSP; P00396; LOCC.
DR Mendel; 5483; ORYsa:cox1.1.
DR InterPro; IPR000885; -.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PR01165; CYCOXIDASE1.
DR PROSITE; PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 64
FT METAL 243
FT METAL 247
FT METAL 292
FT METAL 293
FT METAL 378
FT METAL 380
FT METAL 380
SQ SEQUENCE 524 AA; 57767 MW; F769B9E7106724E0 CRC64;

alignment_scores:
Quality: 66.00 Length: 108
Ratio: 1.245 Gaps: 6
Percent Similarity: 49.074 Percent Identity: 27.778

alignment_block:
US-09-540-234-1 x COX1_ORYSA ..

Align seg 1/1 to: COX1_ORYSA from: 1 to: 524

198 AATGGCGTC...GATGTTATCGAGCCATTCGTGCTGTACTGTGTA 244
:::|||||:::|||||:::|||||:::  ::: |||:::
156 SerGlyValSerSerIleLeuGlySerIleAsnPheIleThrThrIlePh 172
:::|||||:::|||||:::|||||:::  ::: |||:::
245 GCGCTTG.....ATGACTTCCTGTGTACTCCGCTT 276
:::|||||:::|||||:::|||||:::  ::: |||:::
172 eAsnMetArgGlyProGlyMetThrMetHisArgLeuProLeuPheValt 189
:::|||||:::|||||:::|||||:::  ::: |||:::
277 GGAGCCATTGAGTCTGTTTACCGTGCAGCAGCTCGAGACTTCCTGCT 326
||||| ||| ::::::::::|:::|||||::: |||
189 rpSerValLeuValThrAlaPheLeuLeuLeuSer.....LeuPro 203
||||| ||| ::::::::::|:::|||||::: |||
327 GTACTCGCTGCGCAGCATAGTCTGTTTGTG..... 356
||||| ||| ::::::::::|:::|||||::: |||
204 ValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnTh 220
||||| ||| ::::::::::|:::|||||::: |||
357 .....CCTGGTCTGGGACACCTGTGTACTACACTT 390
||||| ||| ::::::::::|:::|||||::: |||
220 rThrPhePheAspProAlaGlyGlyAspProIleLeuTyrGlnHisL 237
||||| ||| ::::::::::|:::|||||::: |||
391 CGTTT.....CCCTTTGAGTGGTACTACTTCTT..... 419
||||| ||| ::::::::::|:::|||||::: |||
237 euPheTrpPheGlyHisProGluValTyrIleLeuLeuProGly 253
||||| ||| ::::::::::|:::|||||::: |||
420 .....AGCCATGTACTC 431
||||| ||| ::::::::::|:::|||||::: |||
254 PheGlyIleIleSerHisIleVal 261
||||| ||| ::::::::::|:::|||||::: |||
seq_name: SwissProt_39:COX1_ARATH
```



```

seq_documentation_block:
ID COX1_ARATH STANDARD; PRT; 527 AA.
AC Q07063;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1 OR COXI.
OS Arabidopsis thaliana (Mouse-ear cress), and
OS Raphanus sativus (Radish).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702, 3726;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=A.thaliana; STRAIN=CV. COLUMBIA;
RX MEDLINE=97141919; PubMed=8988169;
RA Unseld M., Marienfeld J.R., Brandt P., Brennicke A.;
RT "The mitochondrial genome of Arabidopsis thaliana contains 57 genes
RT in 366,924 nucleotides";
RL Nat. Genet. 15:57-61(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=R.sativus;
RX MEDLINE=91330332; PubMed=1651176;
RA Makarov C.A., Apel I.J., Palmer J.D.;
RT "The role of coxi-associated repeated sequences in plant
RT mitochondrial DNA rearrangements and radish cytoplasmic male
RT sterility";
RL Curr. Genet. 19:183-190(1991).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y08502; CAA69795.1; -
DR EMBL; X57692; CAA40873.1; -
DR PIR; S14139; S14139.
DR HSP; P00396; 10CC.
DR InterPro; IPR000883; -.
DR InterPro; IPR002428; -.
DR Pfam; PF00115; COX1; 1.
DR PROSITE; PS01165; CYCOXIDASE1.
DR PROSITE; PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 64 64 IRON (HEME A) (PROBABLE).
FT METAL 243 243 COPPER B (PROBABLE).
FT METAL 247 247 COPPER B (PROBABLE).
FT METAL 292 292 COPPER B (PROBABLE).
FT METAL 293 293 COPPER B (PROBABLE).
FT METAL 378 378 IRON (HEME A3) (PROBABLE).
FT METAL 380 380 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 527 AA; 57996 MW; 69509955882FDE9E CRC64;

alignment_scores:
Quality: 66.00 Length: 108
Ratio: 1.245 Gaps: 6
Percent Similarity: 49.074 Percent Identity: 27.778

alignment_block:
US-09-540-234-1 x COX1_ARATH ..
Align seg 1/1 to: COX1_ARATH from: 1 to: 527
198 AATGGGGTC...GATGTTATCGGAGCCATTCGTGCTGCTTACTGTGTA 244
:::||||| ::::::::::::::::::::: ||| ::|
156 SerGlyValSerSerIleLeuGlySerIleAsnPheIleThrIlePh 172
245 GCGCTTG.....ATGACTTCTCTGCTGCTACTCCGCTT 276
:::| ::| ::| ::|
172 eAsnMetArgGlyProGlyMetThrMetHisArgLeuProLeuPheValT 189
277 GGAGCCATTTGAGTGTGTTTACCGTCGACGACTCGAAGACTTCCTTGCT 326
||||| ||| ::::::::::: ||| ::|
189 rPserValLeuValThrAlaPheLeuLeuLeuSer.....LeuPro 203
327 GTACTCGCTGGCAGCATAGTCTGTTTGTCTG..... 356
||||| ::| ::| ::|
204 ValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnTh 220
357 .....CCTGCTGGGGACACCTGCTGCTACTACAGTACTT 390
||| ||| ::| ::|
220 rThrPhePheAspProAlaGlyGlyGlyAspProIleLeuTyrlGlnHisL 237
391 CGTTT.....CCCTTTCTAGTGTGCTACTTCTT..... 419
||| ||| ::| ::|
237 euPheTrpPheGlyHisProGluValTyrlleLeuLeuProGly 253
420 .....AGCCATGTACTC 431
||||| ::|
254 PheGlyIleIleSerHisIleVal 261

seq_name: SwissProt_39:COX1_OENBE
seq_documentation_block:
ID COX1_OENBE STANDARD; PRT; 527 AA.
AC P08743;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1 OR COI.
OS Oenothera bertiana (Bertero's evening primrose).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Myrtales; Onagraceae; Oenothera.
OX NCBI_TaxID=3950;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. MUNZIIA;
RA Hiesel R., Schobel W., Schuster W., Brennicke A.;
RT "The cytochrome oxidase subunit I and subunit III genes in Oenothera
RT mitochondria are transcribed from identical promoter sequences.";
RL EMBO J. 6:29-34(1987).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

```

CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; MITOCHONDRIAL  
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
CC -|- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL:	X05465:	CAA29025.1;	-	
DR	DR	PIR:	A26170:	ODOB1M.
DR	DR	HSP:	P00396:	1QCC.
DR	DR	InterPro:	IPR000883:	-
DR	DR	InterPro:	IPR002428:	-
DR	DR	Pfam:	PF00115:	COX1; 1.
DR	DR	PRINTS:	PR01165:	CYCOXIDASE1.
DR	DR	PROSITE:	PS00077:	COX1; FALSE_NEG.
DR	DR	Oxidoreductase:	Heme; Copper:	Mitochondrion; Transmembrane;
DR	DR	Respiratory chain:	Inner membrane.	
FT	FT	METAL	64	64
FT	FT	METAL	243	243
FT	FT	METAL	247	247
FT	FT	METAL	292	292
FT	FT	METAL	293	293
FT	FT	METAL	378	378
FT	FT	METAL	380	380
FT	FT	METAL	527 AA;	527242 MW;
SQ	SQ	SEQUENCE	527 AA;	526F5448B5B64ABF CRC64;

```

alignment_scores:
  Quality: 66.00      Length: 109
  Ratio: 1.138       Gaps: 8
  Percent Similarity: 52.294    Percent Identity: 29.358

alignment_block:
  US-09-540-234-1 x COX1_OENBE ..

Align seg 1/1 to: COX1_OENBE from: 1 to: 527

```

198 AATGGCGTC...GAGTTATTCCGAGGCATTCTGTGCTGTACTGTCTGA 244  
       |||||     |||     |||||     |||||     |||     |||||  
 156 SerGlyValSerIleLeuGlySerIleAsnPheIleThrIleSe 172  
       |||||     |||||     |||||     |||||     |||     |||||  
 245 GCGCTTG.....ATGACT.....TCC TTGCTGTACTCCG 273  
       : : : : :     |||||     |||     |||||     :  
 172 rASmEa rGlyLeuGlyMetThrMethHisArgSerProLeuPhe...V 188  
       : : : : :     |||||     |||     |||||     :  
 274 CTTGGAGCCATTGTAGTGC GTTTACCGTGCACGACTCGAAGACTTCCT 323  
       |||||     |||||     |||||     |||||     |||     |||||  
 188 alTrpSerValLeuAlaThrAlaPheProIleLeuLeuSer.....Leu 202  
       |||||     |||||     |||||     |||||     |||     |||||  
 324 GCTGTACTCCTGGCAGCATAGTCTGTTTGGTCG..... 356  
       |||||     |||||     |||||     |||||     |||     |||||  
 203 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 219  
       |||||     |||||     |||||     |||||     |||     |||||  
 357 .....CCTGGCTGGGGACACCTGTGTACTACAGTA 387  
       |||||     |||||     |||||     |||||     |||     |||||  
 219 nThrThrPheSerAspProAlaGlyGlyAspProIleLeuTy rGlnH 236  
       |||||     |||||     |||||     |||||     |||     |||||  
 388 CTTTCGTTT.....CCCTTTCTAGTGGTACTACTTCTT... 419  
       |||     |||     |||     |||     |||     |||||  
 236 lsLeuPheArGPhePheGlyHisProGluValTyrIleLeuIleLeuPro 252

```

420 .....AGCCATGACTC 431
      |||||:
253 GlySerGlyIleIleSerHisIleVal 261
      seq_name: SwissProt_39:COX1_MAIZE
      seq_documentation_block:

```

```

ID AC COX1_MAIZE STANDARD; PRRT; 528 AA.
AC P08742;
AC 01-AUG-1988 (Rel. 08, Created)
AC 01-NOV-1995 (Rel. 32, Last sequence update)
AC 01-NOV-1997 (Rel. 35, Last annotation update)
AC 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
DE COX1 OR COXI.
DE Zea mays (Maize).
DE Mitochondrion.
OG Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OG Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC Andropogoneae; Zea.
OC NCBI_TaxID=4577;
OX [1]
OX SEQUENCE FROM N.A.
RN RN
RN Isaac P.G., Jones V.P., Leaver C.J.;
RA "The maize cytochrome c oxidase subunit I gene: sequence, expression
RA and rearrangement in cytoplasmic male sterile plants.";
RT EMBO J. 4:1617-1623(1985).
RL
RL -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT'S 1-
CC 3 FORM THE CATALYTICAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X02660; CAA26496.1; -.
CC PIR; A22840; ODZM1.
CC HSSP; P00396; LOCC.
CC MaizDB; 69214; -.
DR Mendel; 2201; ZEMa;cox1.1.
DR InterPro; IPR000883; -.
DR InterPro; IPR002428; -.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PR01165; CYCOXIDASE1.
DR PROSITE; PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 64 64 IRON (HEME A) (PROBABLE).
FT FT FT METAL 243 243 COPPER B (PROBABLE).
FT FT FT METAL 247 247 COPPER B (PROBABLE).
FT FT FT METAL 292 292 COPPER B (PROBABLE).
FT FT FT METAL 293 293 COPPER B (PROBABLE).
FT FT FT METAL 378 378 IRON (HEME A3) (PROBABLE).
FT FT FT METAL 380 380 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 528 AA; 58257 MW; 75F895E5568E89E7 CRC64;

```

alignment_scores:		
Quality:	66.00	Length: 108
Ratio:	1.245	Gaps: 6
Percent Similarity:	49.074	Percent Identity: 27.778

```
alignment_block:
US-09-540-234-1 x COX1_MAIZE ..
```

Align seg 1/1 to: COX1\_MAIZE from: 1 to: 528

198 AATGGCGTC...GATGTTATCGGAGCCATTCGTGCTGCTGTTACTGTGTA 244  
 ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||  
 156 SerGlyValSerSerIleLeuGlySerIleAsnPhelThrThrIlePh 172  
 245 GCCTGTC.....ATGACTTCCTTCGCTGCTACTCGCTT 276  
 ::::: ||||| ||||| ||||| |||||  
 172 eAnMetArgGlyProGlyMetThrMetHisArgLeuProLeuPheValt 189  
 277 GGAGCCATTTCAGTGTGTTTACCGTGCAGCAGCTCGAAGACTTCCTTGCT 326  
 ||||| ||||| ::::: ||||| ||||| |||||  
 189 rpSerValLeuValThrAlaPheLeuLeuLeuSer.....LeuPro 203  
 327 GTACTCGTGGCAGCATAGTCTGTTGCTG..... 356  
 ||||| ||||| ::::: ||||| ||||| |||||  
 204 ValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPhaAsnTh 220  
 357 .....CCGTGCTGGGACACCTGTGCTACTAGTACTT 390  
 ||||| ||||| ||||| ||||| |||||  
 220 rThrPhePheAspProAlaGlyGlyGlyAspProIleLeuTyrglnHisL 237  
 391 CGTTT.....CCCTTTGTAGTGTGCTACTACTTCTT..... 419  
 ||||| ||||| ||||| ||||| |||||  
 237 eupHeirPhePheGlyHisProGluValTyrlleLeuLeuProGly 253  
 420 .....AGCCATGTACTC 431  
 254 PheGlyIleIleSerHisIleVal 261

seq\_name: SwissProt\_39:COX1\_SORBI

seq\_documentation\_block:

ID COX1\_SORBI STANDARD; PRT; 530 AA.  
 AC P05502;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
 GN COX1 OR COXI.  
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;  
 OC Andropogoneae; Sorghum.  
 OX NCBI\_TaxID=4558;  
 RN [1]  
 RP SEQUENCE OF 1-31 FROM N.A.  
 RX MEDLINE=87051727; PubMed=3022938;  
 RA Bailey-Serres J., Hanson D.K., Fox T.D., Leaver C.J.;  
 RT "Mitochondrial genome rearrangement leads to extension and relocation  
 of the cytochrome c oxidase subunit I gene in sorghum.";  
 RL Cell 47:567-576(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Hanson D.K., Bailey-Serres J., Leaver C.J.;  
 RL Submitted (MAY-1987) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 AND COPPER B.  
 CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +  
 4 FERRICYTOCHROME C.  
 CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch)  
 CC -----

CC EMBL; M14453; AAA68624.1; -.  
 DR HSP; P00396; LOCC.  
 DR Mendel; 2178; SORBI; cox1;1.  
 DR InterPro; IPR000883; -.  
 DR InterPro; IPR002428; -.  
 DR Pfam; PF00115; COX1; 1.  
 DR PRINTS; P01165; CYCOXIDASE1.  
 DR PROSITE; PS00077; COX1; 1.  
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
 KW Respiratory chain; Inner membrane.  
 FT METAL 64 64 IRON (HEME A) (PROBABLE).  
 FT METAL 243 243 COPPER B (PROBABLE).  
 FT METAL 247 247 COPPER B (PROBABLE).  
 FT METAL 292 292 COPPER B (PROBABLE).  
 FT METAL 293 293 COPPER B (PROBABLE).  
 FT METAL 378 378 IRON (HEME A3) (PROBABLE).  
 FT METAL 380 380 IRON (HEME A) (PROBABLE).  
 SQ SEQUENCE 530 AA; 58522 MW; 18921AE8E7689ABC CRC64;

alignment\_scores:

Quality: 66.00 Length: 108  
 Ratio: 1.245 Gaps: 6  
 Percent Similarity: 49.074 Percent Identity: 27.778

alignment\_block:

US-09-540-234-1 x COX1\_SORBI ..

Align seg 1/1 to: COX1\_SORBI from: 1 to: 530

198 AATGGCGTC...GATGTTATCGGAGCCATTCGTGCTGCTGTTACTGTGTA 244  
 ::|||::| ::::: ||||| ::::: ||||| ::::: |||||  
 156 SerGlyValSerSerIleLeuGlySerIleAsnPhelThrThrIlePh 172  
 245 GCCTGTC.....ATGACTTCCTTCGCTGCTACTCGCTT 276  
 ::::: ||||| ||||| ||||| |||||  
 172 eAnMetArgGlyProGlyMetThrMetHisArgLeuProLeuPheValt 189  
 277 GGAGCCATTTCAGTGTGTTTACCGTGCAGCAGCTCGAAGACTTCCTTGCT 326  
 ||||| ||||| ::::: ||||| ||||| |||||  
 189 rpSerValLeuValThrAlaPheLeuLeuLeuSer.....LeuPro 203  
 327 GTACTCGTGGCAGCATAGTCTGTTGCTG..... 356  
 ||||| ||||| ::::: ||||| ||||| |||||  
 204 ValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPhaAsnTh 220  
 357 .....CCGTGCTGGGACACCTGTGCTACTAGTACTT 390  
 ||||| ||||| ||||| ||||| |||||  
 220 rThrPhePheAspProAlaGlyGlyGlyAspProIleLeuTyrglnHisL 237  
 391 CGTTT.....CCCTTTGTAGTGTGCTACTACTTCTT..... 419  
 ||||| ||||| ||||| ||||| |||||  
 237 eupHeirPhePheGlyHisProGluValTyrlleLeuLeuProGly 253  
 420 .....AGCCATGTACTC 431  
 254 PheGlyIleIleSerHisIleVal 261

seq\_name: SwissProt\_39:SYR\_TREPA

seq\_documentation\_block:

ID SYR\_TREPA STANDARD; PRT; 589 AA.  
 AC O83803;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ARGINYL-TRNA SYNTHETASE (EC 6.1.1.19) (ARGININE--TRNA LIGASE) (ARGRS).  
 GN ARGS OR TP0831.  
 OS Treponema pallidum.





```

72 ProGlyCysLeuAspPheAspProPArgGlnValAsnGluThrTrpTr 88
205 TCGATGTTATCGAGCATTCTGCTGCTTTACTGCTGAGCGCTGATG 254
88 pleuCyS.....:||||:||||:||||:||||:||||:||||:
255 ACTTCCTCTGCTACTCCGCTTGGAGCCATTGAGTGCCTGTTACCG... 301
101 snValValGluIleValGluLeuGlu.....CysAsnProPro 113
302 .....TGACGACTCGA.....AGACTTCCT..... 322
114 ProMetProThrCysSerAsnGlyLeuLysProValArgValProAspPr 130
323 .....TCGCTACTCGCTGCAGCATAGTCTGTTGCTGCCTGCTGTCG 365
130 oGspGlyCysCysTrpHisTrpGluCysAsp.CysTrpCysThrGlyTrp 146
366 GAGCACCT.....GTGCTACTACAGT 386
147 GlyAspProHisPheValThrPheAspGlyLeuTyrTyrSer 160

seq_name: SwissProt_39:CR2_MOUSE
seq_documentation_block:
ID CR2_MOUSE STANDARD; PRT; 1025 AA.
AC P19070;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=90229735; PubMed=2139457;
RA Fingerioth J.D.;
RT *Comparative structure and evolution of murine CR2. The homolog of
RL the human C3d/EBV receptor (CD21).";
RN [2]
RP SEQUENCE OF 12-1025 FROM N.A.
RX MEDLINE=91010789; PubMed=2145366;
RA Molina H., Kinoshita T., Inoue K., Carel J.C., Holers V.M.;
RT "A molecular and immunochemical characterization of mouse CR2.
RN Evidence for a single gene model of mouse complement receptors 1 and
RL J. Immunol. 145:2974-2983(1990).
RN [3]
RP SEQUENCE OF 343-401 AND 991-1025 FROM N.A.
RX MEDLINE=89098890; PubMed=2783485;
RA Fingerioth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
RT "Identification of murine complement receptor type 2.";
RN Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
RN [4]
RP SEQUENCE OF 289-1025 FROM N.A.
RX MEDLINE=89381350; PubMed=2528587;
RA Kurtz C.B., Paul M.S., Aegeerter M., Weis J.J., Weis J.H.;
RT "Murine complement receptor gene family. II. Identification and
RN characterization of the murine homolog (CR2) to human CR2 and its
RL molecular linkage to Crry.";
RL J. Immunol. 143:2058-2067(1989).
CC -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
CC LYMPHOCYTES ACTIVATION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.
CC -1- SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).
CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) DOMAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M81083; AAA37451.1; -
EMBL; M35684; AAA37448.1; -
EMBL; M61132; AAA63295.1; -
EMBL; M35685; AAA37450.1; ALT_SEQ.
EMBL; M29281; AAA37447.1; -
PIR; A43526; A43526.
HSP; P10998; 1VVC.
MGD; MGI:88489; Cr2.
InterPro; IPR000436; -.
Pfam; PF00084; sushi; 14.
Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
Receptor; Sushi.
SIGNAL 1 11
CHAIN 12 1025 COMPLEMENT RECEPTOR TYPE 2.
DOMAIN 12 963 EXTRACELLULAR (POTENTIAL).
TRANSMEM 964 990 POTENTIAL.
DOMAIN 991 1025 CYTOPLASMIC (POTENTIAL).
DOMAIN 13 74
DOMAIN 81 139 SUSHI 1.
DOMAIN 145 203 SUSHI 2.
DOMAIN 206 264 SUSHI 3.
DOMAIN 267 335 SUSHI 4.
DOMAIN 342 399 SUSHI 5.
DOMAIN 401 458 SUSHI 6.
DOMAIN 462 515 SUSHI 7.
DOMAIN 518 586 SUSHI 8.
DOMAIN 593 649 SUSHI 9.
DOMAIN 653 705 SUSHI 10.
DOMAIN 708 769 SUSHI 11.
DOMAIN 777 834 SUSHI 12.
DOMAIN 840 898 SUSHI 13.
DOMAIN 901 959 SUSHI 14.
DOMAIN 14 56 SUSHI 15.
DISULFID 14 56 BY SIMILARITY.
DISULFID 42 73 BY SIMILARITY.
DISULFID 82 124 BY SIMILARITY.
DISULFID 110 138 BY SIMILARITY.
DISULFID 146 189 BY SIMILARITY.
DISULFID 175 202 BY SIMILARITY.
DISULFID 207 248 BY SIMILARITY.
DISULFID 234 263 BY SIMILARITY.
DISULFID 268 317 BY SIMILARITY.
DISULFID 297 334 BY SIMILARITY.
DISULFID 343 385 BY SIMILARITY.
DISULFID 371 398 BY SIMILARITY.
DISULFID 402 445 BY SIMILARITY.
DISULFID 431 458 BY SIMILARITY.
DISULFID 463 501 BY SIMILARITY.
DISULFID 487 514 BY SIMILARITY.
DISULFID 519 568 BY SIMILARITY.
DISULFID 548 585 BY SIMILARITY.
DISULFID 594 636 BY SIMILARITY.
DISULFID 622 649 BY SIMILARITY.
DISULFID 654 689 BY SIMILARITY.
DISULFID 675 704 BY SIMILARITY.
DISULFID 709 752 BY SIMILARITY.
DISULFID 738 769 BY SIMILARITY.
DISULFID 778 820 BY SIMILARITY.
DISULFID 806 833 BY SIMILARITY.
DISULFID 841 884 BY SIMILARITY.
DISULFID 870 897 BY SIMILARITY.
DISULFID 902 945 BY SIMILARITY.
DISULFID 931 958 BY SIMILARITY.
CARBOHYD 77 77 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 113 113 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 276 276 N-LINKED (GLCNAC. .) (POTENTIAL).

```

```

FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 484 484 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 527 527 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 694 694 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 813 813 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 929 929 YGS -> EFR (IN REF. 4).
FT CONFLICT 289 291 S -> T (IN REF. 2).
FT CONFLICT 306 306 S -> A (IN REF. 2).
FT CONFLICT 520 520 MISSING (IN REF. 4).
FT CONFLICT 962 963 MISSING (IN REF. 4).
SQ SEQUENCE 1025 AA; 112994 MW; 19518B9A0273694 CRC64;

alignment_scores:
    Quality: 66.00      Length: 164
    Ratio: 0.917       Gaps: 9
Percent Similarity: 43.902 Percent Identity: 26.220

alignment_block:
US-09-540-234-1/rev x CR2_MOUSE ..
Align seg 1/1 to: CR2_MOUSE from: 1 to: 1025
441 GAACAGACAGAGTACGTACGTGCTAAGAGTAGTACCTACCAAGAGGGAAC 392
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
655 GluProMetArgValHisGly.....LeuProAspAspSerHisIle 668

391 GAAGTACTGTAGTACAGCTGTCCTCCAGACAGGCGAGCAACAGACTAT 342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
668 eLys.LeuValLysArgThrCysGlnAsnGlyTyrGlnLeuThrGlyTyr 684

341 GCTGCCAGCGAGTACAGCAAG..... 321
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
685 Thr.....TyrGluLysCysGlnAsnAlaGluAsnGlyThrTrpPh 698

320 .GAAGTCTTCGAGTCGTCGCGTAAACAGCACTCAAAATGCTCCAAAGCG 272
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
698 eLysLysIleGluValCysThrValIleLeuCysGlnProProLys. 714

271 GAGTACAGCAGGAGTACATCAACGCGTACACAGTAAACAGCACCAAT 222
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
715 ..IleAlaAsnGlyGlyHisThrGlyMetMetAlaLysHisPheLeuTyr 730

221 GGCTCC.....GATAACATCGACGCCATGTCGCGT 193
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
731 GlyAsnGluValSerTyrGluCysAspGluGlyPheTyrLeuLeuGlyG 747

192 TGTTCGTCCTGTAACAACTACGCGTTCACCGACGCTCTCGAGGGGCATGTT 143
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
747 uLysSerLeu.....GlnCysv 753

142 TGAGACACCTCAAAAGGCGAGGACTGTTTCCATTATGCGGTGCCAG 93
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
753 al...AsnAspSerLysGlyHisGlySerTrpSerGlyProProGln 768

92 CTACAGTAGCGTCTCGCGTGTGAGTGTCTTGTCTC..... 54
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
769 CysLeuGlnSerSerProLeuThrHisCysProAspProGluValLysHi 785

53 .....TTCCTTCAGATGACCCACTAGGCGCTCTCGCAT 21
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
785 sGlyTyrLysLeuAsnLysThrHisSerAlaPheSerHis 798

seq_name: SwissProt_39:MA2X_HUMAN
```

```

seq_documentation_block:
ID MA2X_HUMAN STANDARD; PRT; 1139 AA.
AC P49641; Q13754;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ALPHA-MANNOSIDASE IIX (EC 3.2.1.114) (MANNOSYL-OLIGOSACCHARIDE
1,3-1,6-ALPHA-MANNOSIDASE) (MAN IIX).
GN MAN2A2 OR MAN2X.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Melanoma;
RX MEDLINE=96102195; PubMed=8524845;
RA Misago M., Ligo Y.-F., Kudo S., Eto S., Mattei M.-G., Moremen K.W.,
RA Fukuda M.N.;
RT "Molecular cloning and expression of cDNAs encoding human alpha-
mannosidase II and a previously unrecognized alpha-mannosidase IIX
isozyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:11766-11770(1995).
CC -!- FUNCTION: CATALYZES THE FIRST COMMITTED STEP IN THE BIOSYNTHESIS
OF COMPLEX N-GLYCANS. IT CONTROLS CONVERSION OF HIGH MANNOSE TO
COMPLEX N-GLYCANS; THE FINAL HYDROLYTIC STEP IN THE N-GLYCAN
MATURATION PATHWAY.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE TERMINAL 1,3- AND
1,6-LINKED ALPHA-D-MANNOSE RESIDUES IN THE MANNOsyl-
OLIGOSACCHARIDE MAN(5)(GLCNAC)(3).
CC -!- PATHWAY: GLYCOSYLATION.
CC -!- SUBUNIT: HOMODIMER, DISULFIDE LINKED (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI (BY
SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS SEEM TO BE PRODUCED BY
ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D55649; BAA09510.1; -.
DR EMBL: L28821; AAA92022.1; -.
DR MIM: 600988; -.
DR InterPro: IPR000602; -.
DR Pfam: PF01074; Glyco_hydro_38; 1.
KW Hydrolase; Glycosidase; Transmembrane; Glycoprotein; Signal-anchor;
KW Golgi stack; Alternative splicing.
FT DOMAIN 1 5 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 6 26 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 27 796 LUMENAL (POTENTIAL).
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1082 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1120 1120 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLOC 783 796 SRRVDEHEQQVD -> GSGLCFLAEHPKGG (IN
SHORT ISOFORM).
SQ SEQUENCE 1139 AA; 129282 MW; 011CA3089FDC0028 CRC64;

alignment_scores:
    Quality: 66.00      Length: 188
    Ratio: 0.825       Gaps: 12
Percent Similarity: 42.553 Percent Identity: 27.660

alignment_block:
US-09-540-234-1 x MA2X_HUMAN ..
```

Align seg 1/1 to: MA2X\_HUMAN from: 1 to: 1139

```

 9 GGGCCCTTATGTCAGGAGCGCCTAGTGGTCACTCTGAAGGAAAGGAC 58
   ||| ||| ::||| ||| |||||
716 GlyValLeuGlnLeuGlyLeuAspGlyHis..... 727
   59 AAAGACACCTCCAAGCCGACACAGGCTA.....CTGT 90
   ::||| ||||| ::|||:|
728 .ArgThrLeuProSerSerValArgIleTyrLeuHisGlyArgGlnLeuS 744
   91 AGCTGGCAGGCATATGAAACACAGCTCCCTTTTGAGGGTCGTCT 140
   ::||| ||||| ::|||:| ||||| |||||:|
744 erValSerArgHis.....GluAlaPhePro..LeuArgValIleA 757
   141 CAACATGCCCTCGAGAGCGTGGTGAAGCCGTAGTTGTACAAGACTAA 190
   ::||| ::|||:|:|:|:|:|:|:|
757 spSerGlyThrSerAspPheAlaLeuSerAsnArgTyrMetGln..... 771
   191 CAACGCCAATGG.....CGTCG 207
   |||
772 .....ValTrpPheSerGlyLeuThrGlyLeuLeuLysSerIleArgAr 786
   208 ATGTTATCGGAGCCATTCGTGCTGCTTACTGTGTAGCGCTT..... 250
   | ::|||:|:|:|:|:|:|:| |||
786 gValaspGluGluHisGluGlnValaspMetGlnValLeuValTyrG 803
   251 .....GATGACTTCCTTGTCTACTCCGCTTGGAGCCA 283
   ||| ||| ||| |||
803 lyThrArgThrSerLysAspLysSerGlyAlaTyrLeuPheLeuProasp 819
   284 TTTGAGTGTCTTTACCGTGCAGCACTCGAAGACTTCCTTGTGTACT.. 331
   |||:|:| ||| ||| ||||| |||||:|
820 GlyGluAlaSer...ProThrSerProArgSerProCysCysValSe 835
   332 .....CGCTGGCAGCAT.....AGTCTGTTTG 353
   ||| ||||| ::|
835 rLeuLysAlaLeuSerSerGlnArgTrpLeuArgThrMetSerThrPhe 852
   354 CTCCTGTGTCTGGGACACC.....TGTTGACTA 382
   :: ||||| ||| |||||
852 hrArgArgSerGlyPheThrIleCysGlnGlyTrpArgGlyCysLeuTrp 868
   383 CAGTACTTCGTTCCCTTTGTAGTGGTACTACT.....TCTTAGCC 423
   ||| ||||| ||||| |||
869 ThrTyrHisProTyrTrpThrSerGlyThrThrSerThrArgSerTipPr 885
   424 ATGTACTCGTCTC 436
   ||||| ::|
885 ocYstThrSerIle 889
```







Genetics 140, 1353-1366, 1995

A:Title: Complete sequence and gene organization of the mitochondrial genome of the land  
A:Reference number: S59143; MUID:96120351

A:Accession: S59154

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-509 <HAT>

A:Cross-references: EMBL:X83390; NID:g975668; PIDN:CAA58307.1; PID:g975680; GSPDB:GN001.3

C:Genetics:

A:Gene: COI

A:Genome: mitochondrion

A:Genetic code: GGC4

A:Start codon: TTG

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology

C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as-

transmembrane protein

F:1-8/Domain: mitochondrial matrix #status predicted <MM1>

F:8-452/Domain: cytochrome-c oxidase chain I homology <CO1>

F:9-37/Domain: transmembrane #status predicted <TM01>

F:38-46/Domain: intracristal #status predicted <ITC1>

F:47-82/Domain: transmembrane #status predicted <TM02>

F:83-90/Domain: mitochondrial matrix #status predicted <MM2>

F:91-113/Domain: transmembrane #status predicted <TM03>

F:114-136/Domain: intracristal #status predicted <ITC2>

F:137-166/Domain: transmembrane #status predicted <TM04>

F:167-178/Domain: mitochondrial matrix #status predicted <MM3>

F:179-208/Domain: transmembrane #status predicted <TM05>

F:209-223/Domain: intracristal #status predicted <ITC3>

F:224-256/Domain: transmembrane #status predicted <TM06>

F:257-264/Domain: mitochondrial matrix #status predicted <MM4>

F:265-281/Domain: transmembrane #status predicted <TM07>

F:282-293/Domain: intracristal #status predicted <ITC4>

F:294-322/Domain: transmembrane #status predicted <TM08>

F:323-330/Domain: mitochondrial matrix #status predicted <MM5>

F:331-351/Domain: transmembrane #status predicted <TM09>

F:352-365/Domain: intracristal #status predicted <ITC5>

F:366-395/Domain: transmembrane #status predicted <TM10>

F:396-401/Domain: mitochondrial matrix #status predicted <MM6>

F:402-428/Domain: transmembrane #status predicted <TM11>

F:429-441/Domain: intracristal #status predicted <ITC6>

F:442-473/Domain: transmembrane #status predicted <TM12>

F:57, 373/Binding site: heme a iron (His) (axial ligands) #status predicted

F:236, 285, 286/Binding site: copper (His) #status predicted

F:236-240/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted

F:240/Binding site: oxygen (Tyr) #status predicted

F:363/Binding site: magnesium (His) (shared with chain II) #status predicted

F:371/Binding site: heme a3 iron (His) (axial ligand) #status predicted

```

alignment_scores:
  Quality: 79.00      Length: 143
  Ratio: 1.097      Gaps: 8
  Percent Similarity: 50.350      Percent Identity: 27.972

alignment_block:
  US-09-540-234-1 x S59154 ..

Align seg 1/1 to: S59154 from: 1 to: 509

```

```

92  GTCGGCAGCGCCATAATGGAAACAGTCCTCCCTCTTTTGAGGGTCGCTC 141
    |||||::: ||| ::| |||::: |||||
118  AlaGlyThrGly...TriThrValTyrPro.ProLeu.....SerSers 131
    |||||::: ||| ::| |||::: |||||
142  AAACATGCCCCCTCGAGACGTCGGTGAAGCCGCTAGTTTGTACAAGACTAAC 191
    ::: ::| ||| ::: |||||::: :::
131  erLeuAlaHisSerGlyAlaSerValaspLeuAlaIlePheSerLeuHis 147
    ::: ::| ||| ::: |||||::: :::
192  AACGCCAATGGCGTCGATGTCATCGGAGCCATTCCGTGCTGTGTTACTGT 241
    |||||::: ::: |||||::: |||||::: |||||
148  LeuAlaGlyMetSerSerIleLeuGlyAlaIleAsnPheIleThrIle 164
    |||||::: ::: |||||::: |||||::: |||||
242  GTAGCGCTTGG.....ATGACTTCTCTGCTGTACTCCG 273
    : ::::: ||| :::

```

```

156 ePheAsnMetArgSerProGlyMetThrMetGluArgValSerLeuPheV 323
274 CTTGGAGCCATTTCAGTCTGTTTACCGTCGACGACTCGAAGACTTCCTT
:||||| ||| ::|||:: |||::
181 alTrpSerIleLeuValThrValPheLeuLeuLeuSer.....Leu 195
324 GCTGTACTCGCTGGCAGCATAGTCTGTTTCTGCTG..... 356
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
196 ProValLeuAlaGlyAlaIleThrMetLeuThrAspArgAsnPheAs 212
357 .....CCTGGTCTGGGGACACCTGCTGTACTACAGTA 387
||| ||| ||| |||:: |||::
212 nThrSerPhePheAspProAlaGlyGlyAspProIleLeuTrpGlnH 229
388 CTTGCTTT.....CCCTTTCTAGTGGTACTACTTCTT... 419
||| ||| ||| ::|||::|||
229 IsLeuPheTrpPhePheGlyHisProGluValTrIleLeuIleLeuPro 245
420 .....AGCCATGTACTC 431
||| ||| ||| ||| ||| |||
246 GlyPheGlyMetIleSerHisIleLeu 254

seq_name: p1r2:D71390

```

seq\_documentation\_block:  
cytochrome-c oxidase (EC 1.9.3.1) chain I - common lancelet mitochondrion  
C:Species: mitochondrion Branchiostoma lanceolatum (common lancelet)  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 20-Jun-2000  
C:Accession: D71390  
R:Spruyt, N.; Delarbre, C.; Gachelin, G.; Laudet, V.  
Nucleic Acids Res. 26, 3279-3285, 1998  
A:Title: Complete sequence of the amphioxus (Branchiostoma lanceolatum) mitochondrial  
A:Reference number: A71390; MUID:98252550  
A:Accession: D71390  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-515 <SPR>  
A:Cross-references: GB:Y16474; NID:g2929989; PIDN:CAA76249.1; PID:g3292993  
C:Genetics:  
A:Gene: COI

```

alignment_scores:
  Quality: 78.00      Length: 115
             Ratio: 1.300      Gaps: 5
  Percent Similarity: 52.174    Percent Identity: 25.217

alignment_block:
  US-09-540-234-1 x D71390 ..
  Align seq 1/1 to: D71390 from: 1 to: 515

```

[illegible]

```

179 rgValProLeuPheValTrpSerIleTrpValThrAlaTyrLeuLeuLeu 195
      ::::::::::::::::::::
309 CTCGAGAGACTCTCTGCTACTCGTGCAGCATAGTCTGTTGCTG... 356
      ::::: ||| ::::::::::::::::::::
196 Leuser.....LeuProValLeuAlaGlyAlaIleThrMetLeuLeu 210
      ::::::::::::::::::::
357 .....CCTGGCTCTGGGACAC 372
      ||| ||| |
210 rAspArgAsnIleAsnThrThrPhePheAspProSerGlyGlyAspP 227
373 CTGCTGACTACAGTACTTCGTTT.....CCCTTTGTAGTG 407
      ::::: ||::: |||
227 rolleLeuTyrGluHisLeuPheTrpPhePheGlyHisProGluValTyr 243
408 GTACTACTTCTT.....AGCCATGTACTCGTG 434
      ::::::::::::::::::::
244 IleLeuIleLeuProGlyPheGlyIleIleSerHisIleIle 258
seq_name: pir2:T08179

seq_documentation_block:
LRG5 protein - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T08179
R:Gloeckner, G.: Beck, C.F.
submitted to the EMBL Data Library, October 1996
A:Description: Molecular characterization of a gene (LRG5) involved in blue light signal
A:Reference number: Z16399
A:Accession: T08179
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-640 <GLO>
A:Cross-references: EMBL:U73817; NID:g1644369; PID:g1644370
C:Genetics:
A:Gene: LRG5

```

```

alignment_scores:
  Quality: 78.00      Length: 192
  Ratio: 1.099       Gaps: 8
  Percent Similarity: 36.979      Percent Identity: 21.875

alignment_block:
US-09-540-234-1 x T08179 ..

Align seg 1/1 to: T08179 from: 1 to: 640

11 GCCTTATGTATGCAGGAGCGCTAGTGGTCTATCTGAAGAAAGGACAA 60
||| :::: ||::: ||| ||::: |||
427 AlaAlaValProAlaAlaAlaCysSerGlyPheArgGlyGlyArgGlyG 443
||| :::: ||| ::::::::::::::::::::
61 AGACACTCTCAAGCGGACACGGCTACTAGTGCACGGCATATGGA 110
::::: ::::: ::::: ||::: |||
443 yGlyValAlaArg.....ValAlaAlaGlyThrSerArgAla 456
::::: ::::: ::::: ||::: |||
111 AAACAGTCTCCCTCCCTTTTGAGGGTCTCTCAACATGCCCTCGAGAG 160
|||||
456 laGlyValProArgArgLeu..... 462
|||||
161 TCGGTGAAGCCGTAGTTGTGTACAGACTAACACGCAATGGCGTGA... 208
||||| ::::: ||::: |||
463 .....GlnArgArgTrpArgArg 469
208 ..... 208

469 gGlyArgGlyTrpArgArgValArgArgArgGlyAlaGlyArgA 486
209 .....TGTTATCGGACCATTCGTGTGCTGTTTACTGTGTAGCGTGTAT 253
||| ::::: ||| ::::: |||
486 laValCysThrAlaGlyArgCysCysTrpMetThrCysLeuProMetTrp 502
||::: ||::: ||::: ||:::

```

```

254 GACTTCCTTGCTGTACTCGCTTGAGCCATTGAG..... 289
      ::::: ::::: ||::: |||
503 GlySerGlyGlyThrTrpProTrpArgProLeuMetThrProSerArgTh 519
      ::::::::::::::::::::
290 .....TGCTGTTTACCGTCGACGACTCGAAGACTTCCTTGCCTGTA 331
      ::::: ||::: |||
519 rCysAlaCysLeuPro.....ThrProCysCysSerArgT 531
      ::::::::::::::::::::
332 .....CGCTGGCAGCAT 343
      ||| ||| |
531 rPLeuArgArgTrpArgCysGlyTrpAlaProGlyGlyArgTrpArgCys 547
344 AGTCTGTTTGTGCTCGCTGG.....TCTGGGGACACCTGTGTA 381
      ||::: ||::: |||
548 SerLeuCysSerCysTrpArgTrpGlyCysSerGlyArgThrProLeu 564
382 ACAGTACTTCGTT.....TCCCTTTGTAGTGGTACTA 413
      | ::::: |||
564 uProThrTrpValTrpArgArgCysCysArgCysCysArgGlySerA 581
      ::::: ||::: |||
414 CTTCTTAGCCATGTACTCGTGTCTGT 439
      ::::: ||::: |||
581 rgAlaProArgCysAsnTrpValCys 589
seq_name: pir2:T26845

seq_documentation_block:
hypothetical protein Y43F4B.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26845
R:Matthews, L.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z20276
A:Accession: T26845
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-607 <WIL>
A:Cross-references: EMBL:AL021481; PIDN:CAA16336.1; CESP:Y43F4B.7
A:Experimental source: clone Y43F4B
C:Genetics:
A:Gene: CESP:Y43F4B.7
A:Introns: 53/3; 80/3; 110/2; 141/2; 169/3; 229/1; 261/3; 300/2; 320/2; 355/3; 377/1;

alignment_scores:
  Quality: 77.50      Length: 136
  Ratio: 1.062       Gaps: 8
  Percent Similarity: 53.676      Percent Identity: 24.265

alignment_block:
US-09-540-234-1 x T26845 ..

Align seg 1/1 to: T26845 from: 1 to: 607

58 CAAAGACACCTCCAAGCCGACAGCTACTGTAGTGCAGCGGATAAT 107
::::: ::::: ||::: |||
266 GluAsnArgMetGlnSerProHisAlaPheIleSerTrpAsnGlyVal 282
::::: ::::: ||::: |||
108 GGAAACACAGTCCCTGC.....CCTT 127
      : ||::: |||
282 u...AsnSerSerCysLeuValValLeuAlaIlePheSerValThrGly 298
||| |||
128 TTGAGGTCGTCTCAACATGCCCTCGAGAGTCGGTGAAGCCGTAGTT 177
||| |||
298 heTyrGlyTyrLeuSer...LeuGlyAsnAspValLysAspThrAlaThr 313
::: ||::: |||
178 TGTACAAGACTAACACCAATGCGTCGATGTTATCGGAGCCATTCT 227
::: ||::: |||
314 LeuAsnLeuProMetThrProPheTyrGlnThrIleLysLeuMetPhe 330
::: ||::: |||
228 GTGCTGTTTACTGTAGG.....CTTGATGACTTCCTTGTGTA 371
||::: ||::: ||::: ||:::

```

```

alignment_scores:
  Quality: 77.00      Length: 140
  Ratio: 1.132      Gaps: 7
  Percent Similarity: 48.571      Percent Identity: 30.714

alignment_block:
  US-09-540-234-1/rev x A43545      ..
  Align seg 1/1 to: A43545 from: 1 to: 395

412 AGTACCCTACAAAGGGAACGAGTACTGTATGACACAGGTGTCCCCC 363
179 SerSerLysThrSerGlyProProValThrAlaThrGlySerLeuG1 195

```

```
271 GAGTACAGCAAGGAGTCAATCAAGCGCTACACAGTAAACAGCACACGAAT 222
    |||  ::|||  |||  :::::  ::
223 .....ValSerSerArgLysThrSerThrSerThrcl 234
221 GGCTCCGATAACATCG....ACGCCA..... 200
    : |||||:::  ::|||
234 nAspProIleThrThrArgSerProSerGlnGluSerSerGlyMetLeuL 251
199 .....TTGGCGTTG 191
251 euValProMetLeuIleAlaLeuValValLeuAlaLeuValAlaLeu 267
190 TTAGTCTTTACAAACTACGGCTTCACCGACGCTCGAGGGGCATGTTTG 141
    |||||:::  |||  |||||  ::  ::::
268 LeuLeuLeuTrrArglnArgGlnLysArgArg.ThrGlyAlaLeu... 282
140 AGAGGACCCCTC.....AAAGGGCAGGACTGTTTTCATTTAT 103
    |||||  |||||  |||||  ::
283 ....ThrLeuSerGlyGlyLysArgAsnGlyValValAspAlaTrrp 297
102 GCGTGCACGACTACACTA 85
298 AlaGlyProAlaArgVal 303
```

seq\_name: pir1.A47328

seq\_documentation\_block:

N:natural killer cell tumor-recognition protein - human  
A:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumor  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 24-Sep-1999  
C:Accession: A47328  
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.  
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993  
A:Title: A cyclophilin-related protein involved in the function of natural killer cells.  
A:Reference number: A47328; MUID:93133824  
A:Accession: A47328  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1403 <AND>  
A:Cross-references: GB:L04288; NID:g181251; PIDN:AAA35734.1; PID:g181252  
A:Experimental source: NK killer cells from adult blood  
A:Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIP:122800)  
C:Genetics:  
A:Gene: GDB:NKTR  
A:Cross-references: GDB:I37171; OMIM:161565  
A:Map position: 3p23-3p21  
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology  
C:Keywords: alternative splicing; lymphocyte  
F:60-230/Domain: cyclophilin homology <CYP>

```
alignment_scores:
  Quality: 77.00      Length: 76
  Ratio: 1.638
  Percent Similarity: 61.842  Percent Identity: 28.947

alignment_block:
US-09-540-234-1/rev x A47328 ..
Align seg 1/1 to: A47328 from: 1 to: 1403

437 AGACAGCATACATGCTAAGAGTAGTACCAC.....TACAAAGG 397
    |||||:::  :::::  |||
826 LysHisSerSerSerSerGluLysThrLeuHisSerLysThrValLysG1 842
    :|||:::  |||||:::  :::::  ::
396 GAAA...CGAAGTACTGTAGTACACAGGTGTCGCCACGACGACGACAAA 350
    :|||:::  |||||:::  :::::  ::
842 YArgAspArgSerSerCysValArgLysThrSerGluSerArgSerSerL 859
349 CAGACTATGCTGCCAGGATACAGCAAGAGTCTTCAGTCTGCACG 300
    |||||:::  :::::  ::
```

```
859 euAspTyrSerSerAspSer.....GluGlnSerSer 869
299 GTAAACAGCACCTCAATGGCTCCAGCGGAGTACAGCAAGGAAGTCATCA 250
    |||||:::  |||||  |||  ::|||:::  ::
870 ValGlnAlaThrGlnSerAlaGlnGluLysGluLysGlnGlyGlnMetG1 886
249 AGCGTACACAGTAAACAGCACACGAAT 222
886 uArgThrHisAsnLysGlnGluLysAsn 895
seq_name: pir2.T07186
```

seq\_documentation\_block:

C:cytochrome-c oxidase (EC 1.9.3.1) chain I - Chlorella vulgaris mitochondrion (fragmen  
C:Species: mitochondrion Chlorella vulgaris  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: T07186  
R:Watanabe, K.I.; Ehara, M.; Inagaki, Y.; Ohama, T.  
Gene 213, 1-7, 1998  
A:Title: Distinctive origins of group I introns found in the COXI genes of three gree  
A:Reference number: Z15982; MUID:98372069  
A:Accession: T07186  
A>Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-235 <WAT>  
A:Cross-references: EMBL:AB011523; NID:g2943728; PIDN:BAA25069.1; PID:g2943729  
A:Experimental source: strain NIES-227  
C:Genetics:  
A:Gene: COXI  
A:Genome: mitochondrion  
A:Note: intron positions not resolved (incomplete sequence)  
C:Function:  
A:Description: catalyzes the oxidation of four molecules of reduced cytochrome c in t  
Producing two protons  
A:Pathway: oxidative phosphorylation; respiratory chain  
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
C:Keywords: chromoprotein; copper; electron transfer; heme; magnesium; membrane-assoc  
F:109,158,159/Binding site: copper (His) #status Predicted  
F:109-113/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
F:113/Binding site: oxygen (Tyr) #status predicted

```
alignment_scores:
  Quality: 75.50      Length: 115
  Ratio: 1.280
  Percent Similarity: 51.304  Percent Identity: 26.957

alignment_block:
US-09-540-234-1 x T07186 ..
Align seg 1/1 to: T07186 from: 1 to: 235
```

```
174 AGTTTGTACAACTAACACGCCAATGGCTCGATGTTATCGGAGCCAT 223
    :::::  :::  ||:::  :::::  ::|||:::
15 AlaIlePheSerLeuHisLeuAlaGlyAlaSerSerIleMetGlyAla1 31
224 TCGTGTGCTGTTTACTGTGTAGCGCTG.....ATGA 255
    ||:::  :::::  ||:::  :::::
31 eAsnLeuIleThrIlePheAsnMetArgAlaProGlyMetSerMeth 48
256 CTTCCCTTGCTGTAACCGCTTGGACCATTTGAGTCTGTTTACCCTGCA 305
    |||  ::|||  |||  :::::
48 IsArgLeuProLeuPheValTrpSerValIleThrAlaPheLeuLeu 64
306 GCATCGAAGACTTCCTTCTGCTACTCGCTGGCAGCATAGTCTGTTGCT 355
    ||:::  |||  |||||:::  ::|||:::
65 IleLeuSer.....LeuProValLeuAlaGlyGlyIleThrMetLeuLe 79
356 G.....CCTGGTCTGGGA 369
    |||  |||
79 uThrAspArgAsnPheAsnThrThrPheLeuAspProAlaGlyGlyGly 96
370 CACCTGTGTACTACAGTACTCGTTT.....CCCTTTGTA 404
```

```

||||: |||: ||| ||| |||
96 spProfileLeuTyrglnHisLeuPheTrpPhePheGlyHisProGluVal 112
405 GTGCTACTACTCTT.....AGCCATGTACTC 431
||||: |||: ||| ||| |||
113 TyrleuileLeuProGlyPheGlyIleValSerHisValIle 127

```

seq\_name: p1r2:G71442

```

seq_documentation_block:
hypothetical protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Accession: G71442
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weizeneegger, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:98121113
A:Accession: G71442
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-383 <REV>
A:Cross-references: GB:297343; NID:g2245073; PID:g2245094
C:Genetics:
A:Map position: 4COP9-4G3845

```

```

alignment_scores:
Quality: 75.50 Length: 110
Ratio: 1.198 Gaps: 6
Percent Similarity: 57.273 Percent Identity: 27.273

alignment_block:
US-09-540-234-1/rev x G71442
Align seg 1/1 to: G71442 from: 1 to: 383
402 CAAGGGAAACGAAGTACTGTAGTACAGAGTGTCC..... 367
53 GlnGlnAspGlnPheCysValGlyGlnAsnSerProValLeuPh 69
366 .CCAGACCAGCAGCAACACACTATGCTGCCAGCGAGTACAGCAA.... 322
69 eProAspAlaAlaAspProLeuValAlaGlyLysIleMetAsnSerV 86
321 ..GGAAGCTTTCGAGTG.....CTGCACGGTAAA 295
86 alGlyThrArgArgAlaGlyThrLeuSerLysTrpPheHisAsnLys 102
294 CAGCACTAAATGGCTCCAGCGGAGTACAGCAAGGAGTATCAACGCG 245
103 GluHisSerThrSerSerThrThrThrAsnLeuLysLysLysAspLysVa 119
244 TACACAGTAACACAGCAGCAATGGCTCCGATACATCGAGCCATGGC 195
119 lArgValGluAsnAlaHisVal.....His.SerAlaValSer 131
194 GTTGTTAGTCTGTACAACATACCGCTTACCGAGTCTCGAGGGCGAT 145
132 IleAlaAlaLeuAlaAlaGlyLeuAlaSerValThrSerAlaSerAsnCy 148
144 TTTGAGACGACCTCAAAGGGCAGGGA 117
148 s.....LysGlySerGly 152

```

seq\_name: p1r2:T42239

```

seq_documentation_block:
probable phosphoroprotein phosphatase (EC 3.1.3.16) - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42239
R:Sherman, P.M.; Sun, H.; Macke, J.P.; Williams, J.; Smallwood, P.M.; Nathans, J.
Proc. Natl. Acad. Sci. U.S.A. 94, 11639-11644, 1997
A:Title: Identification and characterization of a conserved family of protein serine/
A:Reference number: Z22115; MUID:97471020
A:Accession: T42239
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-707 <SHE>
A:Cross-references: EMBL:AF023454; NID:g2586408; PIDN:AA82794.1; PID:g2586409
C:Keywords: EF hand; phosphoric monoester hydrolase; serine/threonine-specific phosph

```

```

alignment_scores:
Quality: 74.50 Length: 192
Ratio: 0.776 Gaps: 9
Percent Similarity: 50.000 Percent Identity: 23.958

alignment_block:
US-09-540-234-1/rev x T42239
Align seg 1/1 to: T42239 from: 1 to: 707
442 CGAACACAGACAGTACATGCTAAGAGTAGT.....AC 408
9 ArgGlnAsnProSerThrGluLeuLysSerThrArgAlaThrThrTh 25
407 CACTACAAAGGAAACGAAGTACTGTAGTACACAGAGTGTCCCCAGACCAG 358
25 rThrThrSerSerSerGlnArgAsnAsnTyrAsn.....A 37
357 GCACCAACAGACTATGCTGCCAGGAGTACAGCAAGGAGTCTTCAGT 308
37 spAsnAsnGlnAsnThrSerSerSerGlyAsnLysLysGluSerSer 53
307 GCTCAGCGGTAAACAGCACTCAAAAT...GGCTCCAAGCGGAGTACAGCAA 261
54 SerSer.SerLysGlnHisSerSerLysLysSerLysSerLysSerL 70
260 GGAAGTCATCAAGCGCTACA.....CAGTAAACA..... 232
70 yLysAsnArgSerProSerProGlnProGlnLeuThrIleLysSerAla 86
231 .....GCACACGAATGGCTCCGA.....TA 212
87 IleLeuIleGlnLysTrpTyrArgCysGluAlaArgLeuGluAlaAr 103
211 ACATCGACGCCATTGGGCTTGTAGTCTGTGTACAAACTACGCG..... 169
103 gArgAlaThrTrpGlnIlePheThrAlaLeuGluTyrAlaGlyGlu 120
168 .....TTACCCGAGCTCTCGAGGGGCATG 145
120 lnAspGlnLeuLysLeuTyrAspPhePheAlaAspValIleArgAlaMet 136
144 TTTGAGACGCCCTCAAAGGCCA..... 121
137 AlaGluGluAsnGlyLysGlyValGluAsnGlyArgAsnSerProLe 153
120 ...GGAGCTGTTTCCATTATGCGTCCAGCTACAGTACCGTGTCCG 75
153 uMeSerAlaLeuSerHisTyrAlaLysProSerLeuMetAspSerGlu 170
74 GCTTGGAGGTCTTGTTCCTTTCCTTCAGATGACCCACTAGGCGCTCCT 25
170 lyGluThrValLysLysMetLeuGluAspThrSerProThrAsnValasp 186
24 GCATACATAAGGGCCGCTGGACCC 1
::: :::::

```

187 IleAspArgAsnTyrLysGlyPro 194

seq\_name: pir2:F82619

seq\_documentation\_block:  
proton glutamate symport protein XF1937 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: F82619  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: F82619  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-447 <STM>  
A:Cross-references: GB:AE003849; NID:g9107030; PIDN:AAF84739.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Rodrigues, V.; Rosa, A.J.; de M.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1937  
C:Superfamily: C4-dicarboxylate carrier protein

alignment\_scores:  
Quality: 74.00 Length: 104  
Ratio: 1.321 Gaps: 3  
Percent Similarity: 53.846 Percent Identity: 26.923

alignment\_block:

US-09-540-234-1 x F82619 ..

Align seg 1/1 to: F82619 from: 1 to: 447

113 ACAGTCCCTGCGCCTTTTGGGTGCTCTCAACATGCCCTCGACGATC 162  
|||||:|||||: |||||:|||||:|||||: ||| :  
311 ThrLeuProThrAlaLeuArgValAlaAspGluMetLysLeuProLysG1 327  
  
163 GGTGAAGCGGT.....AGTTGTACAGACTAACACGCCAATGCG 203  
:|||||: :|||||: ||| ||| |||||  
327 nile.SerArgPheValLeuThrIleGlyAlaThrAlaAsnGlnAsnGly 343  
  
204 GTCGATGTATCGGACCATTCGTGCTGTTTACTGCTAGCGCTTGAT 253  
:|||||: :|||||: |||||: ||| :  
344 ThrAlaLeuPheGluGlyLeuThrValIlePheLeuAlaGlnPheG1 360  
  
254 GACTTCCTTGCTGCTACTCGCTGGAGCCATTGTAGTGTCTTTACCGTG 303  
:|||||: :|||||: ||| :|||  
360 yValAspLeuSerPheSerGlnAlaMetValMetAla..... 373  
  
304 CAGCACTCGAAGACTCCTTCGTGCTACTCGCTGGCAGCATATGCTGTTTG 353  
|||||: :  
374 .....ValcysIle 376  
  
354 CTGCTGCTGCTGGGACACCTGTGTACTACAGTACTGCTTCCTTCCTTGT 403  
||| |||||:|||||: ||| |||||: |||  
377 LeuGlyGlyIleGlyThrAlaGlyValProSerGlySerLeuProValVa 393

404 ACTGGTACTA 413

|:|||||:

393 lAlaLeuIle 396

seq\_name: pir2:T19551

seq\_documentation\_block:  
mucin-like protein let-653 precursor - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T19551; S69552; S59520  
R:Dobson, R.  
submitted to the EMBL Data Library, May 1996  
A:Reference number: Z19141  
A:Accession: T19551  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-693 <WIL>  
A:Cross-references: EMBL:Z72504; PIDN:CAA96602.1; GSPDB:GN00022; CESP:C29E6.1  
A:Experimental source: clone C29E6  
R:Jones, S.J.M.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: S69552  
A:Accession: S69552  
A:Molecule type: DNA  
A:Residues: 1-693 <JON>  
A:Cross-references: EMBL:X91045; NID:gl199510; PID:e222516; PID:gl199511  
A:Note: this is a revision to the sequence from reference S59520  
R:Jones, S.J.M.; Baillie, D.L.  
Mol. Gen. Genet. 248, 719-726, 1995  
A:Title: Characterization of the let-653 gene in Caenorhabditis elegans.  
A:Reference number: S59520; MUID:96069711  
A:Accession: S59520  
A:Molecule type: DNA  
A:Residues: 1-40, 'T', 42-138, 'POVS', 144, 'MDNNKFRVVSFEFGWK', 162-317, 'I', 319-693 <JOW  
A:Cross-references: EMBL:X91045  
A:Note: this sequence has been revised in reference S69552  
C:Genetics:  
A:Gene: let-653  
A:Map position: 4  
A:Introns: 21/1; 119/1; 212/1; 280/3; 334/2; 374/2; 473/2; 596/3; 658/1  
F;1-23/DNA: signal sequence #status predicted <SIG>  
F;24-693/Product: mucin-like protein let-653 #status predicted <MAT>

alignment\_scores:  
Quality: 74.00 Length: 151  
Ratio: 1.104 Gaps: 6  
Percent Similarity: 44.371 Percent Identity: 24.503

alignment\_block:

US-09-540-234-1/rev x T19551 ..

Align seg 1/1 to: T19551 from: 1 to: 693

439 ACAGACACGAGTACATGGCTAAGAAGTAGTACACTACAAAGGAAACGA 390  
|||||:|||||: |||||:|||||: |||||: |||  
407 ThrGluThrAlaThrThrSerSerThrThrValThr..... 420  
  
389 AGTACTGTAGTACACAGGTGTCCTCCAGACGAGCCAGCAACAGACTATGC 340  
||| ||| :||| :|||  
421 .....ThrGlnLysProThrThrValThrSerThrThrL 433  
  
339 TCCACGCGAGTACAGCAAGAGTCTTCGAGTCTGCACGGTAAACAGCA 290  
|||||:|||||: |||||: |||||: |||||: |||  
433 euProSerThrThrAlaSerThrThrLys..... 443  
  
289 CCAAATGGCTCCAAAGCGGATACAGCAAGGAAGTCATCAAGCGCTACAC 240  
|||  
444 .....Th 444  
  
239 ACTAACAGCACAGCAATGGCTCCGATACATCGACG..... 203  
|:|||||: |||||: |||||: |||||: |||||: |||



```

444 rThrThrSerThrProGlnThrThrThrHisValGlyA 461
202 ..CCATGGCGCTTGTAGCTTGTACAAACTACGGCTTCACCGACGCTC 155
||| :: ::::: ||| ||| |||
461 laProAlaSerSerValAlaSerVal.....AlaHis.AspGlyse 474
154 GAGGGCAGTGTTCAGACGACGCCCTCAAAGGCGAGGACTGTTTCCATT 105
| :: ::::: ::::: ||| ||| |||
474 rThrLeuAlaGlySerProLysValProValIlePheAspIlePheHisA 491
104 ATGCCGTGCCAGCTACAGTACGCCGTGTCGGCTGTGGAGGTGCTTTGTCC 55
:: |||::: ::::: ||| ||| ::::: ||| ||| |||
491 snGlyGlnProValGluAlaValValGlyThrLysIleSerLeuSer 507
54 TTCTCTTCAGATGACCCCATAGGCGCTCTGCATACATAGAGGCCCGTGG 5
||| |||::: ||| |||::: ||| |||::: ||| |||
508 PheArgProHisTyrProIle...ProProGluTyrValAspValArgG 523
4 A 4
|
523 y 523

seq_name: pir2:T02147

seq_documentation_block:
hypothetical protein F8K4.22 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
R:Accession: T02147
R:Vysotskaia, V.S.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu, J.
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; E
submitted to the EMBL data Library, August 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC F8K4 sequen
A:Reference number: Z14574
A:Accession: T02147
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1743 <YYS>
A:Cross-references: EMBL:AC004392; NID:g3282170; PID:g3367535;
C:Genetics:
A:Gene: ATSP:F8K4.22
A:Map position: 1

```

```

284 TTTGAGTGC.....TGTTTACCTGCAGCACTCGAAGACT 318
      |||::| |||::| |||::| |||::| |||
578 ILeGluCysAspTyrPheAlaLeuCysPheGlyCysAlaThr.....Le 592
      |||::| |||::| |||::| |||::| |||
319 TCCTTGTCTACTCGCTGCAGCATAGT.....CTGTTTGCTGCCT 359
      |||::| |||::| |||::| |||::| |||
592 uProGlnGluValArgTyrGluHisAspLysHisMetLeuThrLeuSerI 609
      |||::| |||::| |||::| |||::| |||
360 GGTCTGGGGACACCTGTGTACTACAGTACTCGTTTCCCTTTGT 403
      |||::| |||::| |||::| |||::| |||
609 yrGlyGluGluThrCysIleThrThrTyrTrpCysAspIleCys 623
      |||::| |||::| |||::| |||::| |||
seq_name: pir2:T09071

seq_documentation_block:
SH3 domains-containing protein POSH -.mouse
C:Species: Mus musculus (house mouse)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Sep-2000
C:Accession: T09071
R:Tapon, N.; Nagata, K.; Lamarche, N.; Hall, A.
EMBO J. 17, 1395-1404, 1998
A:Reference number: Z16552; MUID:98151363
A:Accession: T09071
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-892 <TAP>
A:Cross-references: EMBL:AF030131; NID:g3002587; PIDN:AAC40070.1; PID:g3002588
A:Experimental source: cell line Ras-transformed NIH 3T3 cells
A:Note: activates JNK/ SAPK cascade; Rac-binding protein
C:Genetics:
C:Gene: POSH
C:Superfamily: RING finger homology
C:Keywords: signal transduction
F:8-58/Domain: RING finger homology <RRN>

alignment_scores:
      Quality: 74.00      Length: 125
      Ratio: 1.104      Gaps: 7
Percent Similarity: 53.600      Percent Identity: 28.000

alignment_block:
US-09-540-234-1/rev x T09071 ..

Align seg 1/1 to: T09071 from: 1 to: 892

398 GGGAAACGAAGTACTGTAGTACACAGGTGCCCCAGACCAGCAGCAAC 349
      |||::| |||::| |||::| |||::| |||
440 GlySerSerGluGlnIleAlaHisLeuArgProGlnThrArgProSerVa 456
      |||::| |||::| |||::| |||::| |||
348 AGACTATGCTGCCAGCAGGTACAGC.....AAGGAA..... 318
      ||| ||| |||::| |||::| |||
456 lTyrValAlaIleTyrProTyrProArgLysGluAspGluLeuGluL 473
      |||::| |||::| |||::| |||::| |||
317 .....GTCTTCGAGTCTGCACG..... 300
      |||::| |||::| |||::| |||
473 euArgLysGlyGluMetPheLeuValPheGluArgCysGlnAspGlyTyr 489
      |||::| |||::| |||::| |||::| |||
299 GTAACACGACCTCAAAATGGCTCCAAGC.....GGAGTACACCAAGGAAG 256
      |||::| |||::| |||::| |||::| |||
490 TyrLysGlyThrSerMetHisThrSerLysIleGlyValPheProGlyAs 506
      |||::| |||::| |||::| |||::| |||
255 TCATCAGCGGTACACAGTAAACAGCACAGCAATGGCTCCGAT..... 213
      ||| ||| |||::| |||::| |||::| |||
506 nTyrValAlaProValThrArgAlaValThrAsnAlaSerGlnAlaLysV 523
      |||::| |||::| |||::| |||::| |||
212 ..AAATCTGAGCGCATTTGGCGTTGTAGTCTGTACAAACTACGGCTTCA 165
      |||::| |||::| |||::| |||::| |||
523 alSerMetSerThrAlaGlyClnAlaSerArgGlyValThrMetValSer 539
      |||::| |||::| |||::| |||::| |||
164 CCGAGCTCTCGAGGGGATGTTTGGACAGCACCTCAAAAGGGCAGGGACT 115
      |||::| |||::| |||::| |||::| |||

```



C:Accession: T32077  
 R:Graves, T.; Wohlmann, P.; Clarke, K.  
 submitted to the EMBL Data Library, July 1997  
 A:Description: The sequence of C. elegans cosmid F22E5.  
 A:Reference number: Z21119  
 A:Accession: T32077  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-478 <GRA>  
 A:Cross-references: EMBL:AF016681; PIDN:AAB66179.1; GSPDB:GN00020; CESP:F22E5.16  
 A:Experimental source: strain Bristol N2; clone F22E5  
 C:Genetics:  
 A:Gene: CESP:F22E5.16  
 A:Map position: 2  
 A:Introns: 30/2; 182/1; 241/3; 314/3; 419/1; 458/2

alignment\_scores:  
 Quality: 73.50 Length: 122  
 Ratio: 1.131 Gaps: 5  
 Percent Similarity: 53.279 Percent Identity: 27.049

alignment\_block:  
 US-09-540-234-1 x T32077 ..

Align seg 1/1 to: T32077 from: 1 to: 478

```

111 AAACAGTCCCTGCC...TTTTGAGGTGCTCTCAACATGCCCTCGAG 157
      :::::|||||  :::::  |||  :::::  |||
313 GUGUANLeuProGluTyrLeuSerPheGlnSerLeuPro..... 327
      :::::|||||  :::::  |||  :::::  |||
158 ACCTCGGTGAAGCGGTAGTTGTACAGACTAACAACGCCAATGGCGTCG 207
      :::::|||||  :::::  |||  :::::  |||
328 .....AsnPheSerIleTyrGlnAlaAsnAlaLeuLeuPheIleM 341
      :::::|||||  :::::  |||  :::::  |||
208 ATGTTATCGAGGCATT..... 224
      |||||  |||||
341 etValIleThrAlaValIleGlyLeuLeuAlaPheSerPheMet 357
      |||||  |||||
225 CGTGTGCTGTTACTGTGTAGTCCCTTGACTTCCTTGCTGTACTCCGC 274
      |||||  |||||  |||||  |||||  |||||  |||||
358 AlavalLeuTyrAsnIlePheArgMetLeuSerPheMetLysValGlnMe 374
      |||||  |||||  |||||  |||||  |||||  |||||
275 TTGAGCCATTGAGTGCTCTTTACCGTCGACGACTCGAGACTTCCTTG 324
      :::::|||||  :::::  |||  :::::  |||
374 tSerAspThrTyrLysArgHisArgAlaAlaValTyrSerLeuIleA 391
      :::::|||||  :::::  |||  :::::  |||
325 CTGTACTCGTGGCAGTACTGTGTGCTGCTGCTGTGCTGCTGCTGCTG 374
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
391 laGlnPheAlaThrSerIleCysPheLeuPro.....Pro 403
      :::::|||||  :::::  |||  :::::  |||
375 GTGTACTACAGTACTTGTGTTCCCTTGTAGTGTACTACTTCTTAGCCA 424
      :::::|||||  :::::  |||  :::::  |||
404 Ile.....SerLeuValPheValPheLeuLysLeuProAs 416
      :::::|||||  :::::  |||  :::::  |||
425 TGTACTCGTGTCTGTT 440
      :::::|||||  :::::  |||  :::::  |||
416 nProGlnValIleVal 421
      :::::|||||  :::::  |||  :::::  |||

```

seq\_name: pir2:S33642

seq\_documentation\_block:  
 homeotic protein zfh-2 - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 24-Sep-1999  
 C:Accession: S33642; S27817  
 R:Fortini, M.E.; Lai, Z.; Rubin, G.M.  
 Mech. Dev. 34, 113-122, 1991  
 A:Title: The Drosophila zfh-1 and zfh-2 genes encode novel proteins containing both zinc  
 A:Reference number: S33641; MUID:92001539  
 A:Accession: S33642  
 A:Status: preliminary  
 A:Molecule type: mRNA

A:Residues: 1-3005 <FOR>  
 A:Cross-references: EMBL:M63450; NID:g158822; PIDN:AAA29051.1; PID:g158823  
 C:Genetics:  
 A:Gene: zfh-2  
 A:Cross-references: FlyBase:FBgn0004607  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation; zinc finger  
 F:1798-1854/Domain: homeobox homology <HOX1>  
 F:2155-2211/Domain: homeobox homology <HOX2>  
 F:2761-2817/Domain: homeobox homology <HOX3>

alignment\_scores:

Quality: 73.50 Length: 121  
 Ratio: 1.185 Gaps: 3  
 Percent Similarity: 51.240 Percent Identity: 21.488

alignment\_block:

US-09-540-234-1/rev x S33642 ..

Align seg 1/1 to: S33642 from: 1 to: 3005

```

410 TACCAGTACAAAGGGAACGAAGTACTGTAGTACACAGG..... 372
      :::::|||||  :::::  |||  :::::  |||
566 TrpHisTyrLysTyrGlnGluThrLeuGluIleHisMetArgGluLysH1 582
      :::::|||||  :::::  |||  :::::  |||
371 .....TGTCCTCCAGACCCAGGCAGCAAA 350
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
582 sProAspGlyLeuSerAlaCysGlyTyrCysLeuAlaGlyGlnGlnHisP 599
      :::::|||||  :::::  |||  :::::  |||
349 CAGACTATGTCGCCAGCGAG.....TACAGCAAGGAAGTC 315
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
599 roArgLeuAlaArgGlyGluSerTyrSerCysGlyTyrLysProTyrArg 615
      :::::|||||  :::::  |||  :::::  |||
314 TTGAGTGTGTCACGGTAACAGCAGCTCAATGGCTCCAGCGGAGTACA 265
      |||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
616 CysGluIleCysAsnTyrSerThrThrThrLysGlyAsnLeuSerIleH1 632
      :::::|||||  :::::  |||  :::::  |||
264 GCAAGGAAGTCATCAGCGCTACACAGTAAACAGACACCAATGGCTCCG 215
      :::::|||||  :::::  |||  :::::  |||
632 sMetGlnSerAspLysHisLeuAsnAsnMetGlnGluLeuAsnSerSerg 649
      :::::|||||  :::::  |||  :::::  |||
214 ATAACATCGACGCCATTGGCGTTGTAGT.....CTTGTA 180
      :::::|||||  :::::  |||  :::::  |||
649 InAsnMetValAlaAlaAlaAlaAlaAlaValThrGlyLysLeuLeu 665
      :::::|||||  :::::  |||  :::::  |||
179 CAACACTACGCTTCACCGACGCTCTCGAGGGGCATGTTTGAGACGACCTC 130
      :::::|||||  :::::  |||  :::::  |||
666 LeuSerSerSerProGlnValThrAlaAlaCysProSerAsnSerg1 682
      :::::|||||  :::::  |||  :::::  |||
129 AAAAGGCGAGGGA 117
      :::::|||||  |||
682 ySerGlyAlaGly 686
      :::::|||||  |||

```

seq\_name: pir2:D69657

seq\_documentation\_block:  
 methionine adenosyltransferase (EC 2.5.1.6) - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: D69657  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
 C:Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
 A:Ethrich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
 Nature 330, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal  
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
 A:Authors: Schleich, S.; Scofield, R.; Schroeter, F.; Sekiguchi, J.; Sekowska, A.; Se  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya



F;122-144/Domain: intracistal #status predicted <ITC2>  
F;145-174/Domain: transmembrane #status predicted <TM04>  
F;175-186/Domain: mitochondrial matrix status predicted  
F;187-216/Domain: transmembrane #status predicted <TM05>  
F;217-231/Domain: intracistal #status predicted <ITG3>  
F;232-265/Domain: transmembrane #status predicted <TM06>  
F;266-272/Domain: mitochondrial matrix #status predicted  
F;273-289/Domain: transmembrane #status predicted <TM07>  
F;290-301/Domain: intracistal #status predicted <ITC4>  
F;302-330/Domain: transmembrane #status predicted <TM08>  
F;331-338/Domain: mitochondrial matrix #status predicted  
F;339-359/Domain: transmembrane #status predicted <TM09>  
F;360-373/Domain: intracistal #status predicted <ITC5>  
F;374-403/Domain: transmembrane #status predicted <TM10>  
F;404-409/Domain: mitochondrial matrix #status predicted  
F;410-436/Domain: transmembrane #status predicted <TM11>  
F;437-449/Domain: intracistal #status predicted <ITC6>  
F;450-481/Domain: transmembrane #status predicted <TM12>  
F;482-522/Domain: mitochondrial matrix #status predicted  
F;523-531/Binding site: heme a iron (His) #status predicted  
F;532-539/Binding site: copper (His) #status predicted  
F;540-549/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
F;550-559/Binding site: oxygen (Tyr) #status predicted  
F;560-569/Binding site: myristate (Lys) (covalent) #status predicted  
F;570-579/Binding site: magnesium (His) (shared with chain I) #status predicted  
F;580-589/Binding site: heme a3 iron (His) (axial ligand) #status predicted

alignment\_scores:  
Quality: 73.00 Length: 143  
Ratio: 0.986 Gaps: 8  
Percent Similarity: 51.748 Percent Identity: 25.874

alignment\_block:  
US-09-540-234-1 x S62706 ..  
  
Align seg l/1 to: S62706 from: 1 to: 523

seq\_name: pir2:S62706

seq\_documentation\_block:

Cytochrome-c oxidase (EC 1.9.3.1) chain I - green alga (Platymonas subcordiformis) mitochondrion

C;Species: mitochondrion Platymonas subcordiformis  
C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 23-Jul-1999  
C;Accession: S62706  
R;Kessler, U.; Zetsche, K.  
Plant Mol. Biol. 29, 1081-1086, 1995  
A>Title: Physical map and gene organization of the mitochondrial genome from the unicell A.Reference number: S62702; MUID:96145517  
A;Accession: S62706  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-523 <RES>  
A:CROSS-references: EMBL:Z47795; NID:g633584; PIDN:CAA87753.1; PID:g633589  
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, January 1995  
C;Genetics:  
A:Gene: coi  
A:Genome: mitochondrion  
C;Function:  
ns from the cytochrome-c oxidase complex catalyzes the oxidation of four molecules ns from the mitochondrial matrix producing two molecules of water and lowering the conce A;Pathway: oxidative phosphorylation; respiratory chain  
A>Note: chain I directly reduces oxygen on the mitochondrial matrix side of the inner-me C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
C;Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnesium active phosphorylation; oxidoreductase; respiratory chain; transmembrane protein F;1-13/Domain: mitochondrial matrix #status predicted <MM1>  
F;13-460/Domain: cytochrome-c oxidase chain I homology <COI>  
F;14-42/Domain: transmembrane #status predicted <TM01>  
F;43-54/Domain: intracistal #status predicted <ITC1>  
F;55-90/Domain: transmembrane #status predicted <TM02>  
F;91-98/Domain: mitochondrial matrix #status predicted  
F;99-121/Domain: transmembrane #status predicted <TM03>

364 AGACAGCGAGCAAAACAGCATGTGCCTCCGGATGTCAGCAAGAAGTC 315  
|||||::: :::: |||||::: |||||::: |||||::: |||||:::  
414 GlnProGlyThrSerGlySerPrometBroserThirProArggluar 430  
::: |||||::: :::: |||||::: |||||::: |||||::: |||||:::  
314 TTCGAGTGCTGCCGGTAACAACAGCACACTCAATGGCTCCAAGCG..... 272  
::: |||||::: |||||::: |||||::: |||||::: |||||:::  
430 gArGaSerLeuARgllleLeuAsPLeuHISARGLEUARGAlalaSeri 447  
271 ..GACTACGAAGAAGTATCATCAAGCGCTACAGTAGTAACACGACACGA 224  
||| |||||::: |||||::: |||||::: |||||::: |||||:::  
447 leGlualuargMetgluilteuARgARgHSARGSERGLnGlnlncIn 463  
223 ATGGCTCCGATAAACATCAGCCCATTTGGCGTGTAGTCTGTGTACAAACT 174  
::: |||||::: |||||::: |||||::: |||||::: |||||:::  
464 ArgArgHISValThrglySerSerLeuthrr..... 473  
173 ACGGCTTCACCAGCGTCTCGAGGGCATGTTTGAGACGACGACCTCAAAAAGG 124  
|||||::: |||||::: |||||::: |||||::: |||||:::  
474 ..... GlyAspProAspThRG 479  
123 GCAGGAGCTGTTTTCCATTATGCCGTGCAGCTACAGTAGCGTGTCGGG 74  
|||||::: |||||::: |||||::: |||||::: |||||:::  
479 lyGlusERAlaaspPrrothrHISARgAlaARgLEUALAGLUARGPhEARg 495  
73 CTGGAGCGTCTTGTGCTCTTCCTTCAGATGACCCACTAGCGCTCC 26  
|||||::: |||||::: |||||::: |||||::: |||||:::  
496 .... GlyArgPheARgALASer....ArgProGlyARgArgSer 508

seq\_name: pir2:S62706

seq\_documentation\_block:

Cytochrome-c oxidase (EC 1.9.3.1) chain I - green alga (Platymonas subcordiformis) mitoc

C;Species: mitochondrion Platymonas subcordiformis  
C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 23-Jul-1999  
C;Accession: S62706  
R;Kessler, U.; Zetsche, K.  
Plant Mol. Biol. 29, 1081-1086, 1995  
A>Title: Physical map and gene organization of the mitochondrial genome from the unicell A.Reference number: S62702; MUID:96145517  
A;Accession: S62706  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-523 <RES>  
A:CROSS-references: EMBL:Z47795; NID:g633584; PIDN:CAA87753.1; PID:g633589  
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, January 1995  
C;Genetics:  
A:Gene: coi  
A:Genome: mitochondrion  
C;Function:  
ns from the cytochrome-c oxidase complex catalyzes the oxidation of four molecule ns from the mitochondrial matrix producing two molecules of water and lowering the conce A;Pathway: oxidative phosphorylation; respiratory chain  
A>Note: chain I directly reduces oxygen on the mitochondrial matrix side of the inner-me C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
C;Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnesium active phosphorylation; oxidoreductase; respiratory chain; transmembrane protein F;1-13/Domain: mitochondrial matrix #status predicted <MM1>  
F;13-460/Domain: cytochrome-c oxidase chain I homology <COI>  
F;14-42/Domain: transmembrane #status predicted <TM01>  
F;43-54/Domain: intracistal #status predicted <ITC1>  
F;55-90/Domain: transmembrane #status predicted <TM02>  
F;91-98/Domain: mitochondrial matrix #status predicted  
F;99-121/Domain: transmembrane #status predicted <TM03>

alignment\_scores:  
Quality: 73.00 Length: 143  
Ratio: 0.986 Gaps: 8  
Percent Similarity: 51.748 Percent Identity: 25.874

alignment\_block:  
US-09-540-234-1 rev x T49859 ..  
  
Align seg l/1 to: T49859 from: 1 to: 513

364 AGACAGCGAGCAAAACAGCATGTGCCTCCGGATGTCAGCAAGAAGTC 315  
|||||::: :::: |||||::: |||||::: |||||::: |||||:::  
414 GlnProGlyThrSerGlySerPrometBroserThirProArggluar 430  
::: |||||::: :::: |||||::: |||||::: |||||::: |||||:::  
314 TTCGAGTGCTGCCGGTAACAACAGCACACTCAATGGCTCCAAGCG..... 272  
::: |||||::: |||||::: |||||::: |||||::: |||||:::  
430 gArGaSerLeuARgllleLeuAsPLeuHISARGLEUARGAlalaSeri 447  
271 ..GACTACGAAGAAGTATCATCAAGCGCTACAGTAGTAACACGACACGA 224  
||| |||||::: |||||::: |||||::: |||||::: |||||:::  
447 leGlualuargMetgluilteuARgARgHSARGSERGLnGlnlncIn 463  
223 ATGGCTCCGATAAACATCAGCCCATTTGGCGTGTAGTCTGTGTACAAACT 174  
::: |||||::: |||||::: |||||::: |||||::: |||||:::  
464 ArgArgHISValThrglySerSerLeuthrr..... 473  
173 ACGGCTTCACCAGCGTCTCGAGGGCATGTTTGAGACGACGACCTCAAAAAGG 124  
|||||::: |||||::: |||||::: |||||::: |||||:::  
474 ..... GlyAspProAspThRG 479  
123 GCAGGAGCTGTTTTCCATTATGCCGTGCAGCTACAGTAGCGTGTCGGG 74  
|||||::: |||||::: |||||::: |||||::: |||||:::  
479 lyGlusERAlaaspPrrothrHISARgAlaARgLEUALAGLUARGPhEARg 495  
73 CTGGAGCGTCTTGTGCTCTTCCTTCAGATGACCCACTAGCGCTCC 26  
|||||::: |||||::: |||||::: |||||::: |||||:::  
496 .... GlyArgPheARgALASer....ArgProGlyARgArgSer 508

seq\_name: pir2:S62706

seq\_documentation\_block:

Cytochrome-c oxidase (EC 1.9.3.1) chain I - green alga (Platymonas subcordiformis) mito

C;Species: mitochondrion Platymonas subcordiformis  
C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 23-Jul-1999  
C;Accession: S62706  
R;Kessler, U.; Zetsche, K.  
Plant Mol. Biol. 29, 1081-1086, 1995  
A>Title: Physical map and gene organization of the mitochondrial genome from the unicell A.Reference number: S62702; MUID:96145517  
A;Accession: S62706  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-523 <RES>  
A:CROSS-references: EMBL:Z47795; NID:g633584; PIDN:CAA87753.1; PID:g633589  
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, January 1995  
C;Genetics:  
A:Gene: coi  
A:Genome: mitochondrion  
C;Function:  
ns from the cytochrome-c oxidase complex catalyzes the oxidation of four molecule ns from the mitochondrial matrix producing two molecules of water and lowering the conce A;Pathway: oxidative phosphorylation; respiratory chain  
A>Note: chain I directly reduces oxygen on the mitochondrial matrix side of the inner-me C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
C;Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnesium active phosphorylation; oxidoreductase; respiratory chain; transmem

[illegible]

seq\_name: pir2:T20654

seq\_documentation\_block:  
 hypothetical protein F09C8.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T20654  
 R:McMurray, A.  
 submitted to the EMBL Data Library, November 1995  
 A:Reference number: Z19306  
 A:Accession: T20654  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-779 <WIL>  
 A:Cross-references: EMBL:Z68132; PIDN:CAA92220.1; GSPDB:GN00028; CESP:F09C8.2  
 A:Experimental source: Clone F09C8  
 C:Genetics:  
 A:Gene: CESP:F09C8.2  
 A:Map position: X  
 A:Introns: 22/2; 165/1; 266/3; 306/3; 376/1; 462/1; 729/2

alignment\_scores:  
 Quality: 72.50 Length: 128  
 Ratio: 1.133 Gaps: 5  
 Percent Similarity: 50.000 Percent Identity: 21.094

alignment\_block:  
 US-09-540-234-1/rev x T20654 ..

Align seg 1/1 to: T20654 from: 1 to: 779

```

434 CACGAGTACATGGCTAGAGTAGTACCTACCAAGGGAACGAGTAC 385
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 HisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnSerMetLysGlnGlnH 154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
384 TGATGTACACAGGTGTCGCCAGCAGCAGCAGCAGCAGTATGCTGCCA 335
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 sAlaMetHisLeu.....GlnGlnGlnProGlnThrAspTyrLeuSers 169
334 GCGAGTACAGCAGGAGTCTTCGAGTGTGCGAGGTAACAGCAGCTCAA 285
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 er.....GluValCysTyrProSerThrSerSer 178
284 ATGGCTCCAGCGGAGTACAGCAAGGAGTCATCAAGCGCTACACAGTAA 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 TyrGluProSerGlyTyrAsnGluSerGlyTyrCysProGlnAsnSerTh 195
234 A.....CAGCAGCAGCAATGGCTCCGATA 212
195 rIleLeuProAspAspProGlnAsnTyrGlnHisGlyAsnAspGlyAspA 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 ACATCGAGCGCATTTGGC..... 195
212 spGluGluSerPheGlyGluProGluArgTyrGluProMetSerProThr 228
194 .....GTTGTAGCTGTGTACAACTACGCTTCACCGAGCTCTCGAGG 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 PheSerIleHisSerPheAlaProThrThrSerThrAsnThrAlaGluAl 245
150 G.....GCATGTTTGTAGACGACCCCTCA 129
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 aGluAspLeuGluValCysIlePhePheProAla 256

```

seq\_name: pir2:S28066

seq\_documentation\_block:  
 sexual differentiation protein Escp - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 31-Jan-2000  
 C:Accession: S28066; T38926  
 R:Benton, B.K.; Reid, M.S.; Okayama, H.

EMBO J. 12, 135-143, 1993  
 A:Title: A Schizosaccharomyces pombe gene that promotes sexual differentiation encode  
 A:Reference number: S28066; MID:93154314  
 A:Accession: S28066  
 A:Molecule type: mRNA  
 A:Residues: 1-413 <BEN>  
 A:Cross-references: EMBL:X69389; NID:g4949; PIDN:CAA49186.1; PID:g4950  
 R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
 submitted to the EMBL Data Library, February 1996  
 A:Reference number: Z21817  
 A:Accession: T38926  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-413 <PEA>  
 A:Cross-references: EMBL:Z69728; PIDN:CAA93587.1; GSPDB:GN00066; SPDB:SPAC56F8.16  
 A:Experimental source: strain 972h-; cosmid c56F8  
 C:Genetics:  
 A:Gene: escl; SPAC56F8.16  
 A:Map position: 1  
 C:Keywords: DNA binding; nucleus

alignment\_scores:  
 Quality: 72.00 Length: 97  
 Ratio: 1.180 Gaps: 1  
 Percent Similarity: 62.887 Percent Identity: 24.742

alignment\_block:  
 US-09-540-234-1/rev x S28066 ..

Align seg 1/1 to: S28066 from: 1 to: 413

```

398 GGAACAGAGTACTGTAGTACACAGGTGCCCCAGACGAGCAGCAAC 349
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
115 GlySerSerSerTyrValTyrSerValProThrAsnSerThrTh 131
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 AGACTATGCTGCCAGCGAGTACAGCAAGGAGTCTTCGAGTGTGCACGG 299
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 rSerGlnAlaSerAlaLysHisSerAlaValProHisArgSerGlnP 148
298 TAACACAGCTCAATGGCTCCAGCGAGTACAGCAAGGAGTACAGCAAGGAGTCA 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 heGlnSerThrThrLeuThrProSerThrThrAspSerSerSerThrAsp 164
248 GCGCTACACAGTAAACAGCAGCAAGTGGCTCCGATACATCGACGCAT 199
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 ValSerSerSerAspSerValSerThrSerAlaSerSerSerAsnAlaSe 181
198 TGGCGTTGTAGTCTGTACAACTACGGCTTCACCGAGCTCTCGAGGG 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 rAsnThrValSerVal.....ThrSerProAlaSerSerSerAlaThrP 196
148 CATGTTTGCAGACGACCTCAAAAGGCGAGGAGTGTTC 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 roLeuProAsnGlnProSerGlnGlnPheLeuValSer 209

```

seq\_name: pir2:T11317

seq\_documentation\_block:  
 cytochrome-c oxidase (EC 1.9.3.1) chain I - Pedinomonas minor mitochondrion  
 C:Species: mitochondrion Pedinomonas minor  
 C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jan-2000  
 C:Accession: T11317  
 R:Turnel, M.; Lemieux, C.; Burger, G.; Lang, B.F.; Orlis, C.; Plante, I.; Gray, M.W.  
 submitted to the EMBL Data Library, December 1998  
 A:Description: The complete mitochondrial DNA sequences of Nephroselmis olivacea and  
 A:Reference number: Z17261  
 A:Accession: T11317  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-523 <YUR>  
 A:Cross-references: EMBL:AF116775; NID:g4378766; PID:g4378771; PIDN:AAD19669.1  
 C:Genetics:







A:Accession: T07068  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-244 <GOD>  
A:Cross-references: EMBL:X91391; NID:gl1771153; PIDN:CAA62736.1; PID:gi1771154  
A:Experimental source: cultivar Moneymaker; leaf  
C:Genetics:  
A:Gene: LIN7  
C:Superfamily: beta-fructofuranosidase  
C:Keywords: cell wall; glycosidase; hydrolase

alignment\_scores:  
Quality: 71.00 Length: 96  
Ratio: 1.340 Gaps: 3  
Percent Similarity: 55.208 Percent Identity: 26.042

alignment\_block:  
US-09-540-234-1/rev x T07068 ..

Align seg 1/1 to: T07068 from: 1 to: 244

427 ACATGGCTAAGAAGTAGTACCTACTAAAGGGAACGAA...GTACTGTA 381  
||||| |||:|||| |||:||||| |||:|  
48 ThrTrpSerGlySerAlaThrIleLeuProGlyAsnLysProValIleLe 64  
380 GTACACAGGTGCTCCCGACGACGAGCAACAGACTATGCTGCCAGCGA 331  
||||| |||:||||| |||:||||| |||:||||| |||:|  
64 uTyrThrGlyIleIleAspAlaAsnGlnThrGlnValGlnAsnTyrAlaI 81  
330 GTACAGCAAGGAAGCTTCGAGTGTGCGCGGTAAACAGCACTCAATGG 281  
||||| |||:||||| ||| |||:||||| |||:|  
81 leProAlaAsnLeuSerAspProTyrLeuArg.....GluTrp 93  
280 CTCACAGCGGAGTACAGCAAGGAAGTATCAACGCGCTACACAGTAAACG 231  
||||| |||:||||| |||:||||| |||:||||| |||:|  
94 IleLysProAspAsnAsnProLeuIleIleAlaAspGluSerIleAsnLy 110  
230 CACACGAATGGCTCCGATACATCGACGCAATTGGCG..... 194  
||||| |||:||||| |||:||||| |||:||||| |||:|  
110 sThrLysPheArgAspProThrAlaTrpMetGlyLysAspGlyHisT 127  
193 ..TTGTTAGCTTGTACAAACTACGGCTTCACCGACGCT 158  
||||| |||:||||| |||:||||| |||:||||| |||:|  
127 rpArgIleValMetGlySerLeuArgLysHisSerArg 139

seq\_name: pir2.A83867

seq\_documentation\_block:  
homoserine dehydrogenase BH1737 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
C:Accession: A83867  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Fujii, H.  
Nucleic Acids Res 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans  
A:Reference number: A83650; MUID:20263314  
A:Accession: A83867  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-342 <SFO>  
A:Cross-references: GB:AP001513; GB:BA000004; NID:gl0174345; PIDN:BAB05456.1; GSPDB:G  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH1737

alignment\_scores:  
Quality: 72.00 Length: 89  
Ratio: 1.220 Gaps: 2  
Percent Similarity: 66.292 Percent Identity: 30.337

alignment\_block:

US-09-540-234-1/rev x A83867 ..

Align seg 1/1 to: A83867 from: 1 to: 342

```

323 AAGCAAGTCTTCGAGTGTGTCGACGGTAAACAGCACTCAATGGCTCCAAG 274
:::|||||::: |||::: ::|||:::|||||:::::: |||:::
134 GlnGluLeuLysGluLeuAlaAsnGlyValGlnLeuLeuTyrGI 150
:::|||||::: |||::: ::|||:::|||||:::::: |||:::
273 CGGAGTACAGCAAGCAAGTATCATCAAGCGCTACACAGTAAACAGCACAGA 224
|||||::: ::|||::: |||::: |||::: |||::: |||:::
150 uGlyThrValMetSerGlyThrProAlaLeuArgLeuProGluHisThrL 167
:::|||||::: |||::: ::|||:::|||||:::::: |||:::
223 ATGCTCCGATACATCGACCCGATT...GGCGTTGTAGTCT.TGTACA 178
|||||::: |||::: |||::: |||::: |||::: |||:::
167 euLeGlyAsnAsnIleGlnGluValLysGlyIleLeuAsnGlyThrThr 183
|||||::: |||::: |||::: |||::: |||::: |||:::
177 AACTACGGCTTACCGAGCTCTCGAGGGC...ATGTTTGAGAGGACCCCT 131
|||||::: |||::: |||::: |||::: |||::: |||:::
184 AsnTyrIleLeuThrGluMetGluArgGlyLysThrPheGluGlnAlaLe 200
|||||::: |||::: |||::: |||::: |||::: |||:::
130 CAAAGGGCGAGGACTGTTTCCATTATGCGGTGCCAGCTACAGTAGCCG 81
|||||::: |||::: |||::: |||::: |||::: |||:::
200 uGlnAspAlaGlnThrLeuGlyTyrAlaGluAlaAspProThrSerAspV 217
|||||::: |||::: |||::: |||::: |||::: |||:::
80 TGTCGGCGTTGGAGGTG 64
|||||::: |||::: |||::: |||::: |||::: |||:::
217 alGluGlyTyrAspVal 222
|||||::: |||::: |||::: |||::: |||::: |||:::

```

seq\_name: pir2:D96810

seq\_documentation\_block:

hypothetical protein Tlil1.6 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: D96810  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ausen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: D96810  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-530 <STO>  
A:Cross-references: GB:AE005173; NID:g6587860; PIDN:AAF18546.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: Tlil1.6  
A:Map position: 1

alignment\_scores:

Quality:	71.00	Length:	64
Ratio:	1.651	Gaps:	2
Percent Similarity:	67.188	Percent Identity:	32.812

alignment\_block:

US-09-540-234-1/rev x D96810 ..

Align seg 1/1 to: D96810 from: 1 to: 530

```

439 ACAGACAGAGTACATCGCTAAGAGTAGTACCACATAAAGGAAACGA 390
|||||::: |||::: |||::: |||::: |||::: |||:::
65 ThrAspThrSerSerGlnProLysLysSerAspSerGlnLysProGlnGI 81
|||||::: |||::: |||::: |||::: |||::: |||:::
389 AGTACTGTAGTACACAGGTGTCCCGACAGCAGCAGCAACACAGACTATGC 340
::: |||::: |||::: |||::: |||::: |||:::
81 n.....LysProLysProAspGluAsnHisProArgL 92

```

```

339 TGCAGCGAGTACAGCAAGGAAGTCTTCGAGTGTGCTGACGTAACACGCA 290
:::|||||::: |||::: |||::: |||::: |||::: |||:::
92 ySerSerGluSerAlaArgLysSerSerAspSerAlaArg.LysSerIL 108
:::|||||::: |||::: |||::: |||::: |||::: |||:::
289 CTCAATATGGCTCCAGCGGAGTACAGCAAGGAAGTATCATCA 250
|||||::: |||::: |||::: |||::: |||::: |||:::
108 eSerSerGlySerSerArgThrGluSerLysArgPheSer 121
:::|||||::: |||::: |||::: |||::: |||::: |||:::

```

seq\_name: pir2:T45785

seq\_documentation\_block:

hypothetical protein F26013.180 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T45785  
R:Delsen, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: Z23013  
A:Accession: T45785  
A:Molecule type: DNA  
A:Residues: 1-438 <DEL>  
A:Cross-references: EMBL:AL133452  
A:Experimental source: cultivar Columbia; BAC clone F26013  
C:Genetics:  
A:Map position: 3  
A:Introns: 35/1; 67/2; 104/3; 326/3  
A:Note: F26013.180

alignment\_scores:

Quality:	70.50	Length:	166
Ratio:	0.860	Gaps:	9
Percent Similarity:	49.398	Percent Identity:	27.108

alignment\_block:

US-09-540-234-1/rev x T45785 ..

Align seg 1/1 to: T45785 from: 1 to: 438

```

436 GACACAGTATGCTGCCAGCGAGTACAGCAAGG..... 320
:: ||| ::|||:: |||
112 rgProThrArg...SerGlySerThrIleArgProSerAsnIleProThr 127
:: ||| ::|||:: |||
319 ...AAGTCTTCGAGTGTGTCGACGGTAAACAGCACTCAATGGCTCCAGC 273
:::|||||::: |||::: |||::: |||::: |||:::
128 IleArgSerSerValProLysLysThrThrThrGlnIleGlnAl 144
::: |||::: |||::: |||::: |||::: |||:::
272 GGAGTACAGC.....AAGGAAGTCAATCAAGCGCTACACAGTAA 235
|::: ||| |::: |||::: |||::: |||::: |||:::
144 aSerAlaSerValSerSerProLysArgThrValSerArgSerLeuThrP 161
:: ||| ::|||:: |||
234 ACAGCACACGAATGGCTCCGATA...ACATCGACGCCATTGGCGTTGTTA 188
|||||::: |||::: |||::: |||::: |||::: |||:::
161 roSerSerArgLysThrProSerProThrProSerArgIleSer 177
::: |||::: |||::: |||::: |||::: |||:::
187 GTCTTTGTACAAACTACCGCTTCCAGCGCTCGAGGGGCGATGTTTGAGA 138
::: |||::: |||::: |||::: |||::: |||:::
178 ThrThrThrSerThrThrProSerPheLysThrAlaGly..... 190
::: |||::: |||::: |||::: |||::: |||:::
137 CGACCTCAAAGGGCAGGAGCTGTTTTCATATTATGCCGT..... 98
||| |||::: |||::: |||::: |||::: |||:::
191 .AspAlaGlnArgSerArgSerLeuThrProArgAlaLysProGlnIleA 207

```



Percent Similarity:	59.350	Percent Identity:	26.016
Ratio:	0.966	Gaps:	2
Length:	123		

97 ProAlaLeuLeuGlnProArgLeuAlaGlyArgHisGluArgGlnProAs 113

97 ProAlaLeuLeuGI

97 ProAlaLeuLeuGlnProArgLe

```
42 .....CATCTGAAGAA.....AGGACAAGACACCTC 69
   ||| :|||
113 nAlaHisCysGlnGluValAlaIleCysSerAlaThrPheThrProp 130
   ||| :|||
70 CAAGCGGAC.....ACGGCTACTGTAGCTGGCAGGCA 103
   || :|||
130 roProAlaGlyAlaSerSerAsnAlaArgAlaThrValAlaCysProGly 146
   ||| :|||
104 TAATGGAACACAGTCCCTGCCCTTTTGGGGTCGTCCTCAACATGCCCT 153
   ||| :|||
147 CysTrpAsnLysValProThrSerProArgAspTyrArgAsnArgProAl 163
   ||| :|||
154 CGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAATGCC 203
   ||| :|||
163 aProCysSerProArgPro.....TTPA 171
204 GTCGATGTTATCGGACCATTCGTGTCTGTACTGTGTACTGTAGCTTGAT 253
   ||| :|||
171 rGArgSerProArgTrpMetProSerValGlnArgCysArgAla..... 185
254 GACTTCCCTGTGTACTCGCTTGGAGCCATTGTAGTGTCTTTACCG... 301
   ||| :|||
186 .....ProThrCysMetProTh 191
302 .TGCAGCACTCGAAGACTTCCTTGTGTACTCGCTGGCAGCATAGTCTGT 350
   ||| :|||
191 rCysSerAla.....ThrThrCysCysSerThrGlyProHis..... 203
351 TTGCTGCCTGGTCTCGGGACACCTGTGTACTACAGTACTTCTGTTCCCTT 400
   ||| :|||
204 .....LeuAlaGlyLeuIleAspPheTyrAsnAla 213
401 TGTAGTGGT 409
   ||| :|||
214 CysSerGly 216
seq_name: pir2:T20620
seq_documentation_block:
hypochemical protein F08H9.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20620
R:Wild, A.
Submitted to the EMBL Data Library, July 1996
A:Reference number: Z19300
A:Accession: T20620
A>Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-346 <WIL>
A:Cross-references: EMBL:Z77657; PIDN:CA801151.1; GSPDB:GN00023; CESP:F08H9.7
C:Genetics:
A:Experimental source: clone F08H9
A:Gene: CESP:F08H9.7
A:Map position: 5
A:Introns: 17/1; 227/2; 272/2
```

```
alignment_scores:
  Quality: 69.50      Length: 123
  Ratio: 1.053       Gaps: 5
  Percent Similarity: 53.659      Percent Identity: 26.016
alignment_block:
US-09-540-234-1/rev x T20620 ..
Align seg 1/1 to: T20620 from: 1 to: 346
393 ACGAAGTACTGTAGTACACAG.....GTGTCGCCAGACGACGC 356
   ||| :|||
35 ThrSerTyrCysTyrThrLysSerAlaGlnTyrPheSerPheSerGluAl 51
```

```
355 AGCAACACAGACTATGCTCCAGCGAGTACAGCAAGAAAGTCTTTCAGTGC 306
   ||| :|||
51 aAlaAsn.....TyrCysGlnSerIleGlyGlyThrGlnValPheIleI 66
   ||| :|||
305 TGCACGGTAAACAGCACTCA.....AATGGCTCCAGCGGAGTACA 265
   ||| :|||
66 leSerSerThrGluLeuSerTrpLeuThrAspPheThrSerSerSerLeu 82
   ||| :|||
264 GCAAGGAAGTCAATCAGCGCTACACAGTAAACAGCACACGAATGCTCGG 215
   ||| :|||
83 AlaGlnProTrpValAlaThrThrArgAsnThrThrAsn.AsnLysTrpT 99
   ||| :|||
214 ATAACATCAGACCCATTGGCGTTGTTAGTCTTTGTACAAACTACGGCTTCA 165
   ||| :|||
99 yrAsnThrAspGlySerSerProTyrSerPhePheTrpThrThrGlyGlu 115
   ||| :|||
164 CCGAGCTCTCAGAGGGCATGTTTGAGACGACCCCTCAAAAGGCGCAGGACT 115
   ||| :|||
116 ProSerLeuAsnGlyAspCysAlaThrPheTysGlyThrGlyLysAlaGl 132
   ||| :|||
114 GTTTTCCATTATGCGCTGCCAGCTACAGTAGCG.....TGTCGGGCTT 71
   ||| :|||
132 yLeuLysAlaValProCysTyrSerIleGlnProAlaValCysLysGlnM 149
70 GGAGGTGCTTTGTCT 54
   ||| :|||
149 etProAlaLeuCysPro 154
seq_name: pir2:JE0132
seq_documentation_block:
glucose-1-phosphate adenyltransferase (EC 2.7.7.27) chain wml2 - Watermelon
N:Alternate names: ADPglucose pyrophosphorylase chain wml2; ADPglucose synthase chain
C:Species: Citrullus vulgaris S.
C>Date: 10-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 07-May-1999
C:Accession: JE0132
R:Kim, I.J.; Kahng, H.Y.; Chung, W.I.
Biosci. Biotechnol. Biochem. 62, 550-555, 1998
A:Title: Characterization of cDNAs encoding small and large subunits of ADP-glucose p
A:Reference number: JE0131; MUID:9823279
A:Accession: JE0132
A:Molecule type: mRNA
A:Residues: 1-481 <KIM>
C:Superfamily: glucose-1-phosphate adenyltransferase
C:Keywords: nucleotidyltransferase
alignment_scores:
  Quality: 69.50      Length: 105
  Ratio: 1.580       Gaps: 4
  Percent Similarity: 41.905      Percent Identity: 28.571
alignment_block:
US-09-540-234-1/rev x JE0132 ..
Align seg 1/1 to: JE0132 from: 1 to: 481
337 CCAGCGAGTACAGCAAGGAGTCTTCGAGTCTGCACGCTAACAGCACT 288
   ||| :|||
6 ProArgSerThrAlaArgLysLeuThrProGly.....ValAlaTyr 19
287 CAATGGCTCCAGCGGAGTACAGCAAGGAAGTATCATCAAGCGCTACACAG 238
   ||| :|||
19 rSerValLeuMetSerGluLeuSerGlu.....V 29
237 TAAACAGCACGAATGGCTCCGATAACATCAGCGCATGGCGTGTGTTA 188
   ||| :|||
29 alSerSerThrLeuGlnAlaProIlePheGluThrProArgAla..... 43
187 GTCCTGTACAAACTACGGCTTCACCGACGCTCTCGAGGGCGCATGTTGAGA 138
   ||| :|||
43 ..... 43
```

```

137 CGACCCCTCAAAG.....GGCAGGG 118
    |||||:::|
44  .AspProLysLysIleAlaSerIleIleLeuGlyGlyAlaGlyThra 60
    ||
117 ACTGTTTTCATTATGCGTGCCAGCTACAGTAGCCCGTGCCGCTTGA 68
    |||||:::|
60  rgLeuPheProLeuThrSerGlnArgAlaLysProAlaValProIleGly 76
    |||||:::|
67  GGTGCTTTTGTCTT 53
    |||||:::|
77  GlyCysTyrArgLeu 81

```

**THIS PAGE BLANK (USPTO)**



```

EST: *
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

```

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

TITLE The structure and function of the expressed portion of the wheat  
genomes - pre-anthesis spike cDNA library

JOURNAL  
COMMENT

Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818

Email: oanderson@w.usda.gov  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: Stratagene SK primer.

FEATURES  
source

Location/Qualifiers  
1..572  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="WHE1798\_G10\_N20"  
/clone\_lib="Wheat pre-anthesis spike cDNA library"  
/tissue\_type="Spikes before anthesis"  
/dev\_stage="Adult plant"  
/lab\_host="E. coli S04R"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site\_1: EcoRI; Site\_2: XhoI; Plants were grown in the  
greenhouse. Whole spike with awns trimmed, white, green  
and yellow anther were collected and total RNA, and  
poly(A) RNA were prepared, a cDNA library was made, and  
the cDNA clones were in vivo excised to give pBluescript  
phagemids in the T3 Close lab (Choi, Close, Fenton) at  
the University of California, Riverside. Plasmid DNA  
preparations and DNA sequencing were performed in the OD  
Anderson lab (all other authors)."  
BASE COUNT 139 a 134 c 176 g 123 t  
ORIGIN

Query Match 65.1%; Score 289; DB 149; Length 572;  
Best Local Similarity 94.4%; Pred. No. 7e-81;  
Matches 322; Conservative 0; Mismatches 15; Indels 4; Gaps 2;

QY 6 caagggcccttatgtatgc---aggagcgccctagtgggtcattcgaagaaagacaaag 62  
DB 233 CACGGGCCCTTAGTGTTCGGGAGCAGCACCTAGTGGTTCATCTGAAGGAAGACAAAG 292  
QY 63 acactcccaagcgacacgctactgtactgtcgcgcgcataatgaaacagtcctcg 122  
DB 293 ACACCTCCCAAGCGGACGGCGCTACCGTAGCTGGCAGCGCATGATGAAACAGTCCCTG 352  
QY 123 cccttttgagggtcgctctcaaacatgccccctcagacgtcggtgaagccgtagtgttctac 182  
DB 353 -CCTTTTGGGGTGGTCTCAATACGCCCTCGAGACGTCGGTGAAGCCGTAGTTGTAC 411  
QY 183 aagactaacaaacccaatggcgctcgtattcgttcgagccattcgtgtgttactgtg 242  
DB 412 AAGACTTAACAACGCCAATGCGGTCGATGTTGCGGAGCCATTGCTGCTGCTTACTGTG 471  
QY 243 tagcgttgatgacttcccttgcgtgtactcgcgttgagccattcgtgtgttaccgt 302  
DB 472 TAGCGCTTATGACTTCTTCTGCTGTACTCCGCTGGGAGCCATTTGAGTGTGTTACCGT 531  
QY 303 gcagcactcgaagacttcccttgcgtgtactcgcgttgagccattcgtgtgttaccgt 343  
DB 532 GTAGCACTCGAGTAGTTCCTTCTGCTGTACTCGCTGGCAGCAT 572

RESULT 2  
LOCUS BF264372  
DEFINITION HV\_CEA0009D23f Hordeum vulgare seedling green leaf EST library  
VERSION HV\_CEA0009D23f Hordeum vulgare seedling green leaf EST library  
KEYWORDS HV\_CEA0009D23f, mRNA sequence.  
SOURCE BF264372  
ACCESSION BF264372

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	289	65.1	572	149	BF482275	BF482275 WHE1798_G
2	237	53.4	880	146	BF264372	BF264372 HV_CEA000
3	102	23.0	271	165	BE230495	BE230495 99AS608 R
4	96.6	21.8	679	165	BE230496	BE230496 99AS609 R
5	88.4	19.9	458	156	C20067	C20067 C20067 Rice
6	61.4	13.8	433	108	AU163926	AU163926 AU163926
7	61.2	13.8	381	152	BG322351	BG322351 OK-Y2-B15
8	53.2	12.0	482	119	AW677348	AW677348 DGL_5_H04
9	43.4	9.8	578	102	AI834587	AI834587 606070D01
10	43	9.7	417	154	BG458389	BG458389 947045B04
11	43	9.7	750	146	BF267047	BF267047 HV_CEA001
12	38.4	8.6	605	114	AW331079	AW331079 707047A05
13	36.8	8.3	581	226	AQ267877	AQ267877 RPI111-72
14	36.6	8.2	511	137	BE575033	BE575033 946090D12
15	36.6	8.2	578	154	BG458520	BG458520 947045B04
16	36.6	8.2	1268	136	BE515313	BE515313 601235904
17	35.6	8.0	613	24	AI770469	AI770469 606053M11
18	35	7.9	175	103	AI938961	AI938961 scs3ell.Y
19	35	7.9	584	103	AI881530	AI881530 606070D01
20	34.4	7.7	968	219	CNS0062H	AL061700 Drosophila
21	34.2	7.7	829	221	CNS04AOV	AL282208 Tetradon
22	34	7.7	364	113	AW262166	AW262166 xg62b06.X
23	34	7.7	674	247	AZ632025	AZ632025 IM048A04
24	33.8	7.6	863	219	CNS013VO	AL103374 Drosophila
25	33.6	7.6	1101	219	CNS00FEV	AL107029 Drosophila
26	33.4	7.5	386	23	AI685592	AI685592 tu20c07.X
27	33.4	7.5	412	140	BE821365	BE821365 GW7002A4
28	33.4	7.5	523	154	BG508825	BG508825 smc90a12.
29	33.4	7.5	535	148	BF424911	BF424911 sU52e11.Y
30	33.4	7.5	544	120	AW733988	AW733988 sk79b06.Y
31	33.4	7.5	628	242	AZ380524	AZ380524 IM0136D02
32	33.4	7.5	933	137	BE563288	BE563288 601335410
33	33.4	7.5	945	141	BE871340	BE871340 601449457
34	33.2	7.5	225	29	AV327005	AV327005 AV327005
35	33.2	7.5	287	23	AI646741	AI646741 ub65c02.X
36	33.2	7.5	742	247	AZ640083	AZ640083 IM0502K04
37	33	7.4	672	112	AW146599	AW146599 614074C07
38	32.8	7.4	416	225	AQ180148	AQ180148 HS_3200_B
39	32.8	7.4	715	172	BF974202	BF974202 602243835
40	32.8	7.4	887	150	BF575314	BF575314 602134043
41	32.8	7.4	1564	153	BG399947	BG399947 602441492
42	32.6	7.3	539	231	AQ665985	AQ665985 HS_5351_B
43	32.4	7.3	747	240	AZ233398	AZ233398 RPCI-23-B
44	32.4	7.3	840	171	BF964711	BF964711 602267814
45	32.2	7.3	437	225	AQ175373	AQ175373 HS_3212_B

## ALIGNMENTS

RESULT 1  
LOCUS BF482275 572 bp mRNA EST 06-DEC-2000  
DEFINITION WHE1798\_G10\_N20S Wheat pre-anthesis spike cDNA library Triticum  
aestivum cDNA clone WHE1798\_G10\_N20, mRNA sequence.  
ACCESSION BF482275  
VERSION BF482275.1 GI:11565576  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae  
1 (bases 1 to 572)  
REFERENCE Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han  
P.S., Hsiao, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,  
Seaton, C.L. and Tong, J.C.

```

VERSION      BF264372.2  GI:13261453
KEYWORDS     EST.
SOURCE       Hordeum vulgare
ORGANISM     Hordeum vulgare
REFERENCE    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
              ; Triticeae; Hordeum.
AUTHORS      Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
              ,X., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
              ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
              Wood,T.
TITLE        Development of a genetically and physically anchored EST resource
              for barley genomics
JOURNAL      Unpublished (2000)
COMMENT      On Nov 17, 2000 this sequence version replaced gi:11195366.
              Contact: Wing RA
              Clemson University Genomics Institute
              Clemson University
              100 Jordan Hall, Clemson, SC 29634, USA
              Tel: 864 656 7288
              Fax: 864 656 4293
              Email: twing@clemson.edu
              Seq primer: AATTAACCCCTCACTAAAGGG
              High quality sequence stop: 819.

FEATURES             Location/Qualifiers
     source           1..880
                     /organism="Hordeum vulgare"
                     /cultivar="C116155 (Mial3)"
                     /db_xref="taxon:4513"
                     /clone="HV.CEa0009D23f"
                     /clone_lib="Hordeum vulgare seedling green leaf EST
                     library HVCDNA0004 (Erysiphe infected & control)"
                     /tissue_type="seedling green leaf"
                     /lab_host="TJC121"
     note="Vector: lambdaDazAP; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT          195 a 232 c 272 g 177 t 4 others
ORIGIN

Query Match      53.4%; Score 237; DB 146; Length 880;
Best Local Similarity 82.3%; Pred. No. 2.6e-64;
Matches 320; Conservative 0; Mismatches 59; Indels 10; Gaps 4;

Qy 6 caccggcccttatgtatgcaggagcgcttagtggtcatctctgaaggaaagacaaagaca 65
Db 461 CACGAGCCCTTATGTTTTCAGGAGCACCTAGTGGGTCTATCTGAAGGAAAGGACAAAGACA 520

Qy 66 cctcaagccggacacggtacttagtcgacgacgataatggaacacagtcctgccc 125
Db 521 CCTNCAGCCGGACGCGGCGACTAGTGGCAGCGCATGATGTGAANAACAGCCCTGCGCT 580

Qy 126 ttgttaggggtgctcaaacatgcccctcgagacgctggtgaag-----ccgtagttgt 180
Db 581 TTTGAGGCTCTCAGATACGCTCGTTCCTCGAGACGTCGGTGAAGCCGTACCGTAGTTGT 640

Qy 181 acaagactaaacacgaatgctgcgtatgttatcgagaccattcgtgctgttactg 240
Db 641 ACAAGACTAACAAACCAATGGCGTC-ATGTTGTCGGAGCCATTCCTGCTGTTTGTGCT 699

Qy 241 ttagcgtctgtagtactccttgctgactccgcttgagccatttgagtgctgttacc 300
Db 700 TGTATGCTTGATGACTNCTNGCTGTACTCTGCTGGAGCCATTTGAGTGTGTTA-C 758

Qy 301 gtgcagcactgaagactcctcgtgctgactcgtctg-----cagcatagctgttctgctc 357
Db 759 GTGCTACACTGGAAGACTTNCCTACTGCTGCTGCGTGGCAAGCTATCGTCTGTTGCTGG 818

Qy 358 ctggtctgggacacctgtgtactacagt 386
Db 819 GTGGGCTGGGGAAACACTGTGTCCAAACGT 847

```

```

RESULT      3
LOCUS       BE230495
DEFINITION 99AS608 Rice Seedling Lambda ZAPII cDNA Library Oryza sativa cDNA
clone 99AS608, mRNA sequence.
ACCESSION  BE230495
VERSION     BE230495.1  GI:8956580
KEYWORDS   EST.
SOURCE     Oryza sativa.
ORGANISM  Oryza sativa
REFERENCE  1 (bases 1 to 271)
AUTHORS   Lee,M.C., Shin,Y.C., Lee,T.H., Jeong,S.H., Kim,J.K., Eun,M.Y. and
           Namh,B.H.
TITLE     Large-scale Sequencing Analysis of ESTs from Rice Seedling
JOURNAL   Unpublished (1999)
COMMENT   Contact: Eun M.Y.
           Department of Cytogenetics
           National Inst. of Agri. Sci. and Tech, RDA
           Suwon, Kyunggido, Korea
           Tel: 82 331 290 0301
           Fax: 82 331 290 0307
           Email: myeun@sun20.asti.re.kr.

FEATURES             Location/Qualifiers
     source           1..271
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /db_xref="taxon:4530"
                     /clone="99AS608"
                     /clone_lib="Rice Seedling Lambda ZAPII cDNA Library"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector at 5' end with EcoRI and 3' end with Xho I site"
BASE COUNT          66 a 66 c 72 g 66 t 1 others
ORIGIN

Query Match      23.0%; Score 102; DB 165; Length 271;
Best Local Similarity 72.6%; Pred. No. 1.4e-21;
Matches 146; Conservative 0; Mismatches 51; Indels 4; Gaps 1;

Qy 6 caccggcccttatgtatgcaggagcgcttagtggtcatctcgaaggaaagacaaagaca 65
Db 12 CACGAGCTCTTAGTGTTCAGAAACATCTAGTGGACCATCTCAAGAAATGCGCAAGACA 71

Qy 66 cctcaagccggacacggtacttagtcgacgacgataatggaacacagtcctgccc 125
Db 72 GCTCAAGCCGAGCGCGCCACTGTGCGAGCAGCGCATATATGGAAGCACTCCCTGCTT 131

Qy 126 ttgttaggggtgctcaaacatgcccctcgagacgctggtgaagccgtagttgttacaag 185
Db 132 TTGTGAG-----ATCTCGACATGCGCCTTACTTTTGGGCTGAAGCCGTAGTTGTACAA 187

Qy 186 actaacacgccaatgcgctc 206
Db 188 ACTAACGACATCAGNGGCATC 208

RESULT      4
LOCUS       BE230496
DEFINITION 99AS609 Rice Seedling Lambda ZAPII cDNA Library Oryza sativa cDNA
clone 99AS609, mRNA sequence.
ACCESSION  BE230496
VERSION     BE230496.1  GI:8956581
KEYWORDS   EST.
SOURCE     Oryza sativa.
ORGANISM  Oryza sativa
REFERENCE  1 (bases 1 to 271)
AUTHORS   Lee,M.C., Shin,Y.C., Lee,T.H., Jeong,S.H., Kim,J.K., Eun,M.Y. and
           Namh,B.H.
TITLE     Large-scale Sequencing Analysis of ESTs from Rice Seedling
JOURNAL   Unpublished (1999)
COMMENT   Contact: Eun M.Y.
           Department of Cytogenetics
           National Inst. of Agri. Sci. and Tech, RDA
           Suwon, Kyunggido, Korea
           Tel: 82 331 290 0301
           Fax: 82 331 290 0307
           Email: myeun@sun20.asti.re.kr.

FEATURES             Location/Qualifiers
     source           1..271
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /db_xref="taxon:4530"
                     /clone="99AS608"
                     /clone_lib="Rice Seedling Lambda ZAPII cDNA Library"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector at 5' end with EcoRI and 3' end with Xho I site"
BASE COUNT          66 a 66 c 72 g 66 t 1 others
ORIGIN

Query Match      23.0%; Score 102; DB 165; Length 271;
Best Local Similarity 72.6%; Pred. No. 1.4e-21;
Matches 146; Conservative 0; Mismatches 51; Indels 4; Gaps 1;

Qy 6 caccggcccttatgtatgcaggagcgcttagtggtcatctcgaaggaaagacaaagaca 65
Db 12 CACGAGCTCTTAGTGTTCAGAAACATCTAGTGGACCATCTCAAGAAATGCGCAAGACA 71

Qy 66 cctcaagccggacacggtacttagtcgacgacgataatggaacacagtcctgccc 125
Db 72 GCTCAAGCCGAGCGCGCCACTGTGCGAGCAGCGCATATATGGAAGCACTCCCTGCTT 131

Qy 126 ttgttaggggtgctcaaacatgcccctcgagacgctggtgaagccgtagttgttacaag 185
Db 132 TTGTGAG-----ATCTCGACATGCGCCTTACTTTTGGGCTGAAGCCGTAGTTGTACAA 187

Qy 186 actaacacgccaatgcgctc 206
Db 188 ACTAACGACATCAGNGGCATC 208

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
1 (bases 1 to 679)  
Lee,M.C., Shin,Y.C., Lee,T.H., Jeong,S.H., Kim,J.K., Eun,M.Y. and  
Nahm,B.H.  
Large-scale Sequencing Analysis of ESTs from Rice Seedling  
Unpublished (1999)  
Contact: Eun M.Y.  
Department of Cytogenetics  
National Inst. of Agri. Sci. and Tech, RDA  
Suwon, Kyunggido, Korea  
Tel: 82 331 290 0301  
Fax: 82 331 290 0307  
Email: myeunesun20.asti.re.kr.

```

FEATURES
source
Location/Qualifiers
1. .679
/organism="Oryza sativa"
/cultivar="Milyang23"
/db_xref="taxon:4530"
/clone="g9AS609"
/clone_lib="Rice Seedling Lambda ZAPII cDNA Library"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/notes="vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site"

BASE COUNT      179 a      171 c      176 g      150 t      3 others
ORIGIN

Query Match      21.8%; Score 96.6; DB 165; Length 679;
Best Local Similarity 71.1%; Pred. No. 9.4e-20;
Matches 143; Conservative 0; Mismatches 54; Indels 4; Gaps 1;

```

Query Match	21.8%;	Score 96.6;	DB 165;	Length 679;
Best Local Similarity	71.1%;	Pred. No. 9.4e-20;		
Matches 143; Conservative	0;	Mismatches 54;	Indels 4;	Gaps
QY     6	cacggcgcccttatgtatgcaggagcgcctagtgggtcatctgaagaaagacaagaaca	65		
Db     12	CACGAGCTCTTAGTTTCAGAACATCTATTGGACCATCTCAAGAAATGGCAAGACA	71		
QY     66	ctccaagcggacacggtactctagtgtgcacgcgcataatggaaaacagtcacctgcc	125		
Db     72	GCTCCAAGCGGCACGGCGCCACTGTGGCAGCACGGCTTATTGGAAGACACTCCCTGCCIT	131		
QY     126	tttgagggcgctctcaaacaccctcgagcgcgcggtgaagcgcgatgttggtacaag	185		
Db     132	TTGTGAG----ATCTCGGACATGGGCCCTTACTTTGGGGTGAAGCCGTAGTTTGTAAC	187		
QY     186	actaacacgcgaatggcgctc	206		
Db     188	ACTAACGACATCACTGGCATC	208		

RESULT	5
C20067/c	
Locus	EST
DEFINITION	24-OCT-1996 C20067 458 bp mRNA C20067 Rice panicle at ripening stage Oryza sativa cDNA clone E11462_1A, mRNA sequence.
ACCESSION	C20067.1 GI:1632338
VERSION	EST.
KEYWORDS	Oryza sativa.
SOURCE	Oryza sativa
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Eriaristidae; Oryzaceae; Oryza. 1 (bases 1 to 458)
REFERENCE	Sasaki,T.
AUTHORS	Rice cDNA from panicle at ripening stage
TITLE	Unpublished (1996)
JOURNAL	Contact: Takuji Sasaki
COMMENT	National Institute of Agrobiological Resources Rice Genome Research Program 2-1-2 Kannondai,Tsukuba Ibaraki,

```

Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@agr.affrc.go.jp.
      Location/Qualifiers
    1. .458
       /organism="Oryza sativa"
       /strain="Nipponbare"
       /db_xref="taxon:4530"
       /clone="E11462_1A"
       /clone_lib="Rice panicle at ripening stage"
       /dev_stage="ripening stage"
       /note="Organ: panicle; Rice cDNA from panicle at ripening
stage"
FEATURES
source

```

BASE COUNT	124 a	113 c	103 g	117 t	1 others
ORIGIN					
Query Match	19.9%	Score 88.4	DB 156	Length 458	
Best Local Similarity	68.5%	Pred. No. 3.4e-17			
Matches 170	Conservative 0	Mismatches 66	Indels 12	Gaps 3	
Qy 6	cacgggcccattgtatgcaggagcgccctagtgggctcatctgaaggaaaggacaaaagaca 65				
Db 345	CACGAGCTCTTAGGTTTGCAGAAACATCTAGTGGACCATCTCAAGAAATATGGCAAGACA 286				
Qy 66	cctccaaagc-----ggacacggctactgtagctggcagcgccataatggaacaacatcc 119				
Db 285	GCTCAAAGCGACGACGAGGCGCGGCCACTCTGGCAGCGCGGCATAATGGAAGCACATCCC 226				
Qy 120	ctgcccttttaggggtcgctcctcaaacatgcctctcgagacgtcggtgaacccgtagttg 179				
Db 225	TGCCCTTTTGTGAG----ATCTCGGACATGGCCCCCTTACTTTTGGGGGTGAAGCCGTAGTTTG 170				
Qy 180	tacaagactaaacagccaatggcgctc--gatgttatcggagccattcgtgtgctgttta 237				
Db 169	TACAACACTAACGACATCAGTGGCATCATGTTTTCCTCTAGCCATGTTGAGTGCATTATA 110				
Qy 238	ctgtgtag 245				
Db 109	CTGTGTTG 102				

RESULT	6
AU163926/c	
LOCUS	433 bp mRNA EST 14-NOV-2000
DEFINITION	Rice panicle at ripening stage Oryza sativa cDNA clone E11462, mRNA sequence.
ACCESSION	AU163926
VERSION	AU163926.1 GI:11171718
KEYWORDS	EST.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa.
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoidae; Oryzaceae; Oryza.
AUTHORS	1 (bases 1 to 433)
TITLE	Sasaki, T. and Yamamoto, K.
JOURNAL	Rice cDNA from panicle at ripening stage (2000)
COMMENT	Unpublished (2000) Contact: Takui Sasaki

2-1-2 Kannondai, Tsukuba  
Ibaraki,  
Japan 305  
Tel: 0298-38-7441  
Fax: 0298-38-7468  
Email: tsasaki@abr.affrc.go.jp  
PROJECT = 'RGP',  
ELL462\_42.

FEATURES                      Location/Qualifiers  
SOURCE                      1. .433



[illegible]

RESULT	9
LOCUS	A1834587/c
DEFINITION	A1834587 578 bp mRNA EST 606070D01.x1 606 - Ear tissue cDNA library from Schmidt lab Zea mays cDNA, mRNA sequence.
ACCESSION	A1834587
VERSION	A1834587.1 GI:5468796
KEYWORDS	EST.
SOURCE	Zea mays.
ORGANISM	Zea mays. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 578) Walbot,V.
REFERENCE	Maize ESTs from various cDNA libraries sequenced at Stanford University
AUTHORS	Unpublished (1999)
TITLE	Contact: Walbot V
JOURNAL	Department of Biological Sciences Stanford University
COMMENT	855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 606070 row: D column: 01.

FEATURES	source
location/Qualifiers	
1. 578	
/organism="zea mays"	
/cultivar="Ohio43"	
/db_xref="taxon:4577"	
/clone_lib="606 - Ear tissue cDNA library from Schmidt lab"	
/tissue_type="mixed"	
/dev_stage="ear length from 0.5 cm - 2.0 cm"	
/lab_host="XLOLR (Stratagene)"	
/note="Organ: Immature ear; Vector: pBK-CMV; Site_1: EcoRI Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt lab"	
BASE COUNT	149 a 152 c 124 g 151 t 2 others
ORIGIN	

Query Match	9.8%	Score 43.4;	DB 102;	Length 578;
Best Local Similarity	57.5%;	Pred. No. 0.0071;		
Matches 157;	Conservative	0;	Mismatches 106;	Indels 10;
				Gaps 4;

[illegible]

Qy	195	gcaaatggcgctga-----tggtatcggagcaccattcgctgtctgtttactgtgtgtagcgctt	250
Db	301	ATGGACGGGGGCGCCGTGTGTTTCCCACTCATGTTCAGTGAATGTTTACCGCGTATGACTT	242
Qy	251	gatgacttccttgctgactcgccttgtagcca	283
Db	241	TATTAGTTGCTGGGCACACACTGGCAGTAGCTA	209

  

RESULT	10
BG458389/c	
LOCUS	947045B04.x2 947 - 2 week shoot from Barkan lab Zea mays CDNA, mRNA
DEFINITION	EST 19-MAR-2001
ACCESSION	BG458389
VERSION	947045B04.1
KEYWORDS	GI:13381609
SOURCE	EST.
ORGANISM	Zea mays
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
AUTHORS	clade; Panicoideae; Andropogoneae; Zea.
TITLE	1 (bases 1 to 417)
JOURNAL	Walbot, V.
COMMENT	Maize ESTs from various cDNA libraries sequenced at Stanford University
	Unpublished (1999)
	Contact: Walbot V
	Department of Biological Sciences
	Stanford University
	855 California Ave, Palo Alto, CA 94304, USA
	Tel: 650 723 2227
	Fax: 650 725 8221
	Email: walbot@stanford.edu
	Plate: 947045 row: B column: 04.

## FEATURES

```

location/Vendurraireis
1. .417
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="947 - 2 week shoot from Barkan lab"
/tissue_type="leaf and stem, including leaf base"
/dev_stage="2 week old seedling (3 leaves)"
/lab_host="X11-Blue"
/note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-);
Site_1: EcoRI; Site_2: XhoI; Directionally cloned using
Stratagene's Unizap XR cDNA cloning kit with the 5' end
at the EcoRI site. The library represents 8 x 105
independent recombinant phage. The plants were greenhouse
grown"

```

BASE COUNT	100 a	118 c	85 g	113 t	1 others
ORIGIN	grown.				

Query Match	9.7%	Score 43;	DB 154;	Length 417;
Best Local Similarity	53.7%;	Pred. No. 0.0087;		
Matches 159;	Conservative	0;	Mismatches 126;	Indels 11;
				Gaps 3;

Qy	20	tatgcaggagcgctagtgggt---catctgaaggaagacacacacccctccaagccg	76
Db	310	TTTTCAGAAAGCGTCTGTGTGATCAGCAGCTCGGGAAGCGGAACACAGCTCGAAGCCG	251
Qy	77	gacacggctactgtactgacgcacgcataatggaataacagtcaccttcttgggggtc	136
Db	250	GACGCAGCCGCTGTATCCGCACAGCTGTATTGTAG-CACTATGCTCTGGACGAGGTCTC	192
Qy	137	gtctcaaacatgcccctcgagacgcggtgaagccgtagtttgtcaagactaacacgc	196
Db	191	AGAAATATTATCTCCTAGCAGACTTTGGTGAACCCATAGCTTTTGAAAAAANAANA	132
Qy	197	caatggcgctcga-----tggtatcggagccattcgtgtcgtgttactgtgtagcgt	249
Db	131	CATGGACGGGGCGCTCTGTGTTTCCAGTCATGTCAGTGATGTTTACCGGTATGACT	72

```
Qy 250 tgaatcttcctgtgactcgtactccgcttgagccattgagtcgtgtttaccgtgca 305
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 TTATTAGTTGTCGGGCACACTGGCAGCTAGCTAGATTGCTGCTGCTGTTGTGA 16

RESULT 11
BF267047 750 bp mRNA EST 09-MAR-2001
LOCUS HV_CEA0016M17f Hordeum vulgare seedling green leaf EST library
DEFINITION HVCDNA0004 [Erysiphe infected & control] Hordeum vulgare cDNA clone.
            HV_CEA0016M17f, mRNA sequence.
ACCESSION BF267047
VERSION BF267047.2 GI:13263140
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 750)
AUTHORS Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
, T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and
Wood, T.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics
JOURNAL Unpublished (2000)
COMMENT On Nov 17, 2000 this sequence version replaced gi:11198042.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACCCCTCAATAAGGG
High quality sequence stop: 745.
FEATURES
    source
        1..750
        /organism="Hordeum vulgare"
        /cultivar="Cili16155 (Mial13)"
        /db_xref="taxon:4513"
        /clone="HV_CEA0016M17f"
        /clone_lib="Hordeum vulgare seedling green leaf EST
        library HVCDNA0004 (Erysiphe infected & control)"
        /tissue_type="seedling green leaf"
        /lab_hosts="TJC121"
        /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 177 a 218 c 227 g 128 t
ORIGIN

Query Match 9.7%; Score 43; DB 146; Length 750;
Best Local Similarity 65.5%; Pred. No. 0.01;
Matches 78; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

Qy 8 cgggcccattatgtatgcaggagcgcctagtgggtcatctatggaagaaagacacacc 67
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 CCGAGCCCTTATGTTGAGGACCCCTA-TGGGGCTTCTGTAGGAAAGACAAAACACC 543
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 68 tccaagccggacacggtactgtactggtgcacggcataatggaacacagtcctgcct 126
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 544 TCGAACCCGGACCCCGTAACTGATCTCGGTCGCCGTTATTGAAGAAAACACCCCGT 602
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AW331079 605 bp mRNA EST 31-JAN-2000
LOCUS AW331079
DEFINITION AW7047A05.x1 707 - Mixed adult tissues from walbot lab (SK) zea
mays cDNA, mRNA sequence.
ACCESSION AW331079
VERSION AW331079.1 GI:6827436
```

```
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
ciade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 605)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 707047 row: A column: 05.
FEATURES
    source
        1..605
        /organism="Zea mays"
        /cultivar="W23"
        /db_xref="taxon:4577"
        /clone_lib="707 - Mixed adult tissues from Walbot lab (SK)"
        /tissue_type="tassel, kernel, silk, husk, root, leaf"
        /dev_stage="adult"
        /lab_hosts="DH10B"
        /note="organ: tassel, kernel, silk, husk, root, leaf;
        vector: pGAD10; Site_1: EcoRI; cDNA library from fully
        differentiated maize tissues from an active Mutator
        plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk,
        husk, root, leaf). Unidirectionally cloned."
BASE COUNT 154 a 155 c 191 g 105 t
ORIGIN

Query Match 8.6%; Score 38.4; DB 114; Length 605;
Best Local Similarity 58.3%; Pred. No. 0.28;
Matches 105; Conservative 0; Mismatches 71; Indels 4; Gaps 2;

Qy 20 tatcgaggagcgctagtgggt---catctgaagaaagaaagacacacctcccaagcgc 76
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 TTGCGAAGCGCTGCTGTGATCAGCAGCTCGGGAACGCGAAAGACAGCTCCAAAGCGC 481
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 77 gacacggctactgtagctggcacggcataatggaaaacagtcctccttttgagggtc 136
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 GACGCGCGCTGTACCGGCACAGCTTGATTTGAG-CACTCTGGCGGAGGTCTC 540
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 137 gtcctcaaacatgccctcgagacgctcgatgaagccgtagttgtacagactaacacgc 196
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 AGATATATTGCTCTCTAGGAGACTTTGGTGAGCCATAGCTTTTGTACACCACTAACCGC 600
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AQ267877/c 581 bp DNA GSS 27-APR-1999
LOCUS RPC111-72E14.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-72E14,
DNA sequence.
ACCESSION AQ267877
VERSION AQ267877.1 GI:3795481
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 581)
AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: RPC111-72E14.TK
```



Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac\_end\_search.html  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
1. .581  
/organism="Homo sapiens"  
/db\_xref="GDB:7527373"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-72E14"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"  
BASE COUNT 260 a 129 c 81 g 111 t  
ORIGIN  
Query Match 8.3%; Score 36.8; DB 226; Length 581;  
Best Local Similarity 48.1%; Pred. No. 0.88;  
Matches 104; Conservative 0; Mismatches 112; Indels 0; Gaps 0;  
QY 200 tggcgtgattatcgaggccattcgtgtgctgttactgtgtgtagcgttgatgacttc 259  
Db 217 TGTGTCTATGATATTTTCAGCAGCTGTTCTAGTATTTTCTGTAGAGGTATTTCACTTC 158  
QY 260 cttgctgtactcgttgaggcattgctgtgtgtttaccgtgcagcactcgaagactt 319  
Db 157 CTGTGGTTAGGTATATTCCTAAGCATTTTATTTATTTATTTGACGCTGTTAAAGG 98  
QY 320 ccttgctgtactcgtggcagcatagctgtgtgtcgtcgtgctgtgggacacctgtgta 379  
Db 97 GGTGTGATCTTTCAGTCTGATCTTCAGCTGCTGCTGCTGTGTGTATAGGAGACTACTGA 38  
QY 380 ctacagtactctgtccctcttgtagtgactact 415  
Db 37 TTGCGTACATTAATTTGTGTCTGGAACGTTACT 2

RESULT 14  
LOCUS BE575033 511 bp mRNA EST 15-AUG-2000  
DEFINITION 946090D12.y1 946 - tassal primordium prepared by Schmidt lab zea  
mays cDNA, mRNA sequence.  
ACCESSION BE575033  
VERSION BE575033.1 GI:9824936  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 511)  
AUTHORS Walbot,V.  
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University  
JOURNAL Unpublished (1999)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 946090 row: D column: 12.  
Location/Qualifiers  
1. .511

FEATURES  
source

/organism="Zea mays"  
/cultivar="OH43"  
/db\_xref="taxon:4577"  
/clone\_lib="946 - tassal primordium prepared by Schmidt lab"  
/tissue\_type="tassels"  
/dev\_stage="just after the transition from vegetative to inflorescence development"  
/lab\_host="XLOLR"  
/note="Organ: tassels; Vector: HybriZAP; Site\_1: EcoRI; Site\_2: XhoI; George Chuck dissected immature tassels between 1mm and 3mm. Sharon Stanfield prepared the cDNA library in HybriZAP. Sample insert size range was 350 bp to 3 Kb with a 1 Kb average."  
BASE COUNT 130 a 132 c 169 g 80 t  
ORIGIN

Query Match 8.2%; Score 36.6; DB 137; Length 511;  
Best Local Similarity 70.3%; Pred. No. 0.99;  
Matches 64; Conservative 0; Mismatches 24; Indels 3; Gaps 1;  
QY 20 tatcgaggagcctagtgtggt---catctgaaggaaagacaaaagacacctcagccg 76  
Db 398 TTTGCAAGAGCGTCTGTGTGATCAGCAGCTCGGGAAGAAAGCAGACAGCTCGAAGCG 457  
QY 77 gacacgctactagtgtgacgcgcataat 107  
Db 458 GACCGAGCGCTGTTACCGGCACAGCTTGAT 488

RESULT 15  
LOCUS BG458520 578 bp mRNA EST 19-MAR-2001  
DEFINITION 947045B04.y1 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA sequence.  
ACCESSION BG458520  
VERSION BG458520.1 GI:13381845  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 578)  
AUTHORS Walbot,V.  
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University  
JOURNAL Unpublished (1999)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 947045 row: B column: 04.  
Location/Qualifiers  
1. .578

FEATURES  
source  
1. .578  
/organism="Zea mays"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone\_lib="947 - 2 week shoot from Barkan lab"  
/tissue\_type="leaf and stem, including leaf base"  
/dev\_stage="2 week old seedling (3 leaves)"  
/lab\_host="XLI-Blue"  
/note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-);



```

RESULT 18
A1938961/c
LOCUS
DEFINITION
  sc3ell.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
  Gm-c1016-1125 5' similar to TR:022152 O22152 F4L23.30 PROTEIN. ;
  mRNA sequence.
ACCESSION
  A1938961
VERSION
  A1938961.1 GI:5677831
KEYWORDS
  EST.
SOURCE
  soybean.
ORGANISM
  Glycine max
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
  1 (bases 1 to 175)
REFERENCE
  1 Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
    A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
    Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
    Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
    R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
    R., Waterston,R. and Willson,R.
  Public Soybean EST Project
  Unpublished (1999)
  Contact: Shoemaker R/Public Soybean EST Project
  Public Soybean EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
  Tel.: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  This clone is available through: Genome Systems, Inc. 4633 World
  Parkway Circle St. Louis, Missouri 63134 For further information
  call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
  427-3324 or contact: clones@genomesystems.com or
  info@genomesystems.com web site: www.genomesystems.com
  Trace considered overall poor quality
  Possible reversed clone: similarity on wrong strand
  High quality sequence stop: 1.
FEATURES
  source
    1..175
    /organism="Glycine max"
    /db_xref="taxon:3847"
    /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-1125"
    /clone_lib="Gm-c1016"
    /tissue.type="immature flowers of field grown plants"
    /lab_host="XL10-Gold"
    /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
    XhoI; This cDNA library was constructed from mRNA isolated
    from immature flowers of field grown plants. The cDNA
    library was prepared using the Stratagene pBluescript II
    XR library construction kit. Complementary DNA was
    synthesized from mRNA using a primer consisting of a poly
    (dT) sequence with a XhoI restriction site. EcoRI adapters
    were ligated to the blunt-ended cDNA fragments followed by
    XhoI digestion. The cDNA fragments were directionally
    cloned into the EcoRI-XhoI restriction site of the
    pBluescript vector. The ligated cDNA fragments were
    transformed into XL10-Gold host cells. This library was
    constructed by Dr. Randy Shoemaker and Dr. John
    Erpelting."
BASE COUNT
  54 a 45 c 43 g 33 t
ORIGIN
  Query Match 7.9%; Score 35; DB 103; Length 175;
  Best Local Similarity 57.9%; Pred. NO. 2.4;
  Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 269 ctccgcttgagccatttgagctgtttacogtcagcactcgaagactctctctgtgt 328
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107 cttctctgtgactatcaggattgttgctgtttgtctctctgtctctctttatgg 48
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 329 actcgtgagcagcatgctgttctgtctgctgtgtctgtgtgtgtgtgtgtgtgt 375

```

```

Db 47 ACTGTTATGACATAAGTACGCGCTCTCTCTCTCGGAGGCGCTG 1
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 19
A1881530
LOCUS
DEFINITION
  606070D01.y1 606 - Ear tissue cDNA library from Schmidt lab Zea
  mays cDNA, mRNA sequence.
ACCESSION
  A1881530
VERSION
  A1881530.1 GI:5555579
KEYWORDS
  EST.
SOURCE
  Zea mays.
ORGANISM
  Zea mays
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
  clade; Panicoidae; Andropogoneae; Zea.
  1 (bases 1 to 584)
REFERENCE
  1 Walbot,V.
  Maize ESTs from various cDNA libraries sequenced at Stanford
  University
  Unpublished (1999)
  Contact: Walbot V
  Department of Biological Sciences
  Stanford University
  855 California Ave, Palo Alto, CA 94304, USA
  Tel: 650 723 2227
  Fax: 650 725 8221
  Email: walbot@stanford.edu
  Plate: 606070 row: D column: 01.
FEATURES
  source
    1..584
    /organism="Zea mays"
    /cultivar="Ohio43"
    /db_xref="taxon:4577"
    /clone_lib="606 - Ear tissue cDNA library from Schmidt
    lab"
    /tissue.type="mixed"
    /dev_stage="ear length from 0.5 cm - 2.0 cm"
    /lab_host="XL0LR (Stratagene)"
    /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
    ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
    lab"
BASE COUNT
  143 a 163 c 181 g 97 t
ORIGIN
  Query Match 7.9%; Score 35; DB 103; Length 584;
  Best Local Similarity 69.2%; Pred. No. 3.3;
  Matches 63; Conservative 0; Mismatches 25; Indels 3; Gaps 1;
QY 20 tatgcaggagccctagtgggt---catctgaagaaagacacacacacacacgcg 76
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 TTTGCAGAGACCTCTGGTGGATCAGCAGCTCGGAAACGCGAAGACACAGCTCGAAGCG 541
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 77 gacacgctactgtagctgacgacgacataat 107
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 542 GACCGAGCGCTGGTACCGGCACAGCTTGAT 572
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 20
CNS0062H
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence TEF3 end of BAC #
  BACR13D18 of RPCI-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL061700
VERSION
  AL061700.1 GI:4943900
KEYWORDS
  GSS.
SOURCE
  fruit fly.
ORGANISM
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```

TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis									
JOURNAL	Unpublished									
REFERENCE	2 (bases 1 to 829)									
AUTHORS	Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.									
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence									
JOURNAL	Unpublished									
REFERENCE	3 (bases 1 to 829)									
AUTHORS	Genoscope.									
TITLE	Direct Submission									
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases									
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/Tetraodon">http://www.genoscope.cns.fr/Tetraodon</a> .									
FEATURES	<p>Location/Qualifiers</p> <p>1..829</p> <p>/organism="Tetraodon nigroviridis"</p> <p>/db_xref="taxon:99883"</p> <p>/clone="096C15"</p> <p>/clone_lib="G"</p> <p>/note="Genoscope sequence ID : COBG096AB08SP1-end : PUC-ori"</p>									
BASE COUNT	240 a	186 c	174 g	226 t	3 others					
ORIGIN										
Query Match	7.7%; Score 34.2; DB 221; Length 829;									
Best Local Similarity	56.8%; Pred. No. 6.5;									
Matches	63; Conservative	0;	Mismatches	48;	Indels	0;	Gaps	0;		
Qy	221	cattcgtgctgttctactgtgtagcgttgatgactctctgtgtactcgcgttgag	280							
Db	556	CAGTCGTTTTTCAGPATGTCTTTAGTGTGTGATGTTCTTTGGCGCCCTAGCTGGAG	497							
Qy	281	ccatttagtgcgtttaccgtgcagcactcgaagactctctgtctact	331							
Db	496	ACAGGCAGGACTGTTCTCTGTCTGTATGTGACCGACTCGCTGACT	446							
RESULT	22									
AW262166/c										
LOCUS	AW262166	364 bp	mrna	EST	28-DEC-1999					
DEFINITION	xq62b06.x1 NCI_CGAP_HN9 Homo sapiens cDNA clone IMAGE:2755187 3' similar to contains element LTR5 repetitive element ;, mRNA sequence.									
ACCESSION	AW262166									
VERSION	AW262166.1 GI:5638982									
KEYWORDS	EST.									
SOURCE	human.									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
AUTHORS	1 (bases 1 to 364)									
TITLE	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .									
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index									
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgaps-r@mail.nih.gov">cgaps-r@mail.nih.gov</a> Tissue Procurement: Edward Shillitoe Ph.D., Silvio Gutkind Ph.D., Chidchanok Leethanakul D.D.S., Michael Emmert-Buck M.D. Ph.D. CDNA Library Preparation: David B. Krizman, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://www-bio.llnl.gov/bbrp/image/image.html">www-bio.llnl.gov/bbrp/image/image.html</a> Possible reversed clone: polvt not found									



library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

```
FEATURES
  source
    1. .863
      /organism="Drosophila melanogaster"
      /plasmid="pBelOBAC11"
      /db_xref="taxon:7227"
      /clone_lib="DrosBAC"
      /clone="BACN10C13"
      /note="end : SP6"
BASE COUNT      86 a  167 c  186 g  259 t  165 others
ORIGIN

Query Match      7.6%; Score 33.8; DB 219; Length 863;
Best Local Similarity 35.6%; Pred. No. 8.9;
Matches 78; Conservative 33; Mismatches 108; Indels 0; Gaps 0;

QY 109 gaaacagtcctccctgttttgaggggtcgctctcaaacatgcccctcgagacgctggtgaa 168
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 564 GANTGCAGNCTATNTTTCATGCGTGGTGTGTAATAGTATTTTCGCTTCGTTTTCGTMK 623
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 169 gccgtagtgttacaagaactaaacccaatggcgcgtcgatgttatcgagccattcgtg 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 624 GATGTTMTTNTTKCMWTKMKTMMTMTGKGGGGTCCMTGCGTGMWGMKTKTMMWTK 683
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 229 tgcgttactgtgtagcgttgatgactctcctgtgtactcgcgttgagccatttga 288
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 MTMTTGTGKMGTKKGGKGTGTTTTMTTNTTKCCGTTMTTTCCTGTTCTTTCTTTG 743
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 289 gtgctgttaccgtgcagcaatcgaagactcctctgtg 327
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 CTKCTCMTTTCMGCGGTTGCTTSGBMMTKYGTGGGTG 782
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 25
CNS00FEV      1101 bp  DNA      GSS      03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION      BACR31J08 of RPCI-98 library from Drosophila melanogaster (fruit
                fly), genomic survey sequence.
ACCESSION      AL070299.1  GI:4950527
VERSION      GSS.
KEYWORDS      fruit fly.
SOURCE      Drosophila melanogaster
ORGANISM      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 1101)
AUTHORS      Genoscope.
JOURNAL      Direct Submission
TITLE      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
COMMENT      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                - Web : www.genoscope.cns.fr)
                Determination of this BAC-end sequence was carried out as part of a
                collaboration with the Berkeley Drosophila Genome Project (BDGP).
                The BDGP is constructing a physical map of the Drosophila
                melanogaster genome using these BACs. For further information
                please see http://www.fruitfly.org The BDGP Drosophila
                melanogaster BAC library was prepared by Kazutoyo Osoegawa and
                Aaron Mamoser in Pieter de Jong's laboratory in the Department of
                Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
                NY. The library is named RPCI-98 and was constructed by partial
                EcoRI digestion of Drosophila DNA provided by the BDGP from the
                isogenic strain y2; cn bw sp, the same strain used for the BDGP's
                P1 and EST libraries. A more detailed description of the library
                and how to order individual BAC clones, the entire library, or
                filters for hybridization from the BACPAC Resource Center can be
                found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
```

```
FEATURES
  source
    1. .1101
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone_lib="RPCI-98"
      /clone="BACR31J08"
      /note="end : T7"
BASE COUNT      260 a  200 c  206 g  217 t  218 others
ORIGIN

Query Match      7.6%; Score 33.6; DB 219; Length 1101;
Best Local Similarity 26.5%; Pred. No. 11;
Matches 54; Conservative 61; Mismatches 89; Indels 0; Gaps 0;

QY 189 acaaacgccaatggcgtcgatgttatcgaggcaattcgctgtgttactgttagcgc 248
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 896 AAMHGKTYCCRKGGYKGBDYGCCTAYASARVCCCTGCTAGTGTTCAAARSWATASCSY 955
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 249 ttgatgactctcgtgtactcgcgttgagccatttgatgctgtttaccgtgcagca 308
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 956 TTKTGTGTTTHWVGSGCKADAWCCGATTKMDTKTYBCTAATAGCCACACTGARGGCT 1015
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 309 ctggaactcctctgtgtactcgtcgagcagcatagctgttctgtgcctgtgctgtgg 368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1016 GRGSGACAKNNSMSYGBKRTAMCCTGGCGMGCMGCDYGGASCGTTTYGYCSBGCVSC 1075
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 369 acactgtgtactacagtaactcg 392
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1076 MSAYGYGYKCTCCTCTKGSACKG 1099
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 26
AI685592/c    386 bp  mRNA      EST      15-DEC-1999
LOCUS      tu20c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
DEFINITION      similar to contains PTR5.t3 MSRI repetitive element ;, mRNA
                sequence.
ACCESSION      AI685592
VERSION      AI685592.1  GI:4896886
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 386)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
                Email: cgaps-r@mail.nih.gov
                Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
                Emmert-Buck, M.D., Ph.D.
                CDNA Library Preparation: M. Bento Soares, Ph.D.
                CDNA Library Arrayed by: Greg Lennon, Ph.D.
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E.E. Consortium/LLNL at:
                www-bio.llnl.gov/bbrp/image/image.html
                Insert Length: 1143 Std Error: 0.00
                Seq primer: -40UP from Gibco
                High quality sequence stop: 380.
FEATURES
  source
    1. .386
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:2251596"
      /clone_lib="NCI_CGAP_Pr28"
      /sex="male"
      /dev_stage="adult"
      /lab_host="DH10B"
      /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
```





```

/db_xref="taxon:3847"
/clone:"GENOME SYSTEMS CLONE ID: Gm-cl073-336"
/clone_lib="Gm-cl073"
/tissue_type="seedlings induced for symptoms of SDS
(Sudden Death Syndrome) disease"
/dev_stage="2-3 weeks old"
/lab_host="DH10B"
/notes="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from mRNA isolated
from 2-3 week old seedlings that were induced for symptoms
of SDS (Sudden Death Syndrome) disease by the
translocation of culture filtrate of Fusarium solani f.
sp. glycines (Plant Cell Report 18:375-380). Cultivar
Williams 82 is susceptible to the disease SDS. Plant
tissue (expanded leaves, folded leaves, and new shoots)
were collected at 1, 6, 24, and 48 hrs. after inoculation
and their mRNA pooled equally for cDNA construction. The
library was prepared using the Stratagene pBluescript II
SK(+) library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(GT) sequence with an XhoI restriction site. EcoRI
adaptors were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA insert is protected
from XhoI digestion via methylation during first strand
synthesis. The cDNA fragments were directionally cloned
into the EcoRI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
E.coli ElectroMax DH10B host cells. Plants were inoculated
by Shuxian Li (Glen Hartman lab, University of Illinois).
Library was constructed by Reena Philip and Steve Clough
(Lila Vodkin lab, University of Illinois)."
BASE COUNT      143 a 120 c 131 g 129 t
ORIGIN

```

```

Query Match      7.5%; Score 33.4; DB 154; Length 523;
Best Local Similarity 55.7%; Pred. No. 10;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 269 ctccgcttgagccatttgagctgtttaccgtgcagcactcgaagactctctgtgt 328
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 CTTCTCTATGACTGATTCAGGATTATTGGCTTTGATCTCTGATCTCTTCCTTTTAA 345

Qy 329 actcgtggcagcatagctgttgcctgtgctgggacacctgtgtactac 383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 ACTGGTTATAGCAGAGGTACGGCGCTCCCTCTTCGCGAGGCGCTGTTCACAAC 290

```

```

RESULT 29
BF424911/c 535 bp mRNA EST 28-NOV-2000
LOCUS
DEFINITION
Gm-cl069-y1 Gm-cl069 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl069-621 5' similar to TR:Q9XFBI Q9XFBI YABBY3. ;, mRNA
sequence.
ACCESSION BF424911.1 GI:11412900
VERSION
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 535)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
CONTACT: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project

```

```

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 419.

```

## FEATURES

```

Location/Qualifiers
source
1..535
/organism="Glycine max"
/db_xref="taxon:3847"
/clone:"GENOME SYSTEMS CLONE ID: Gm-cl069-621"
/clone_lib="Gm-cl069"
/tissue_type="Degenerating cotyledons, 9-10 day old
etiolated seedling"
/lab_host="DH10B"
/notes="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from mRNA isolated
from degenerating cotyledons of 9-10 day old etiolated
seedlings for the cultivar Williams. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(GT) sequence with a XhoI restriction site. EcoRI
adaptors were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments
were transformed into DH10B host cells (GibcoBRL). This
library was constructed in the laboratory of Dr. Randy
Shoemaker."
BASE COUNT      139 a 128 c 133 g 134 t 1 others
ORIGIN

```

```

Query Match      7.5%; Score 33.4; DB 148; Length 535;
Best Local Similarity 55.7%; Pred. No. 10;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 269 ctccgcttgagccatttgagctgtttaccgtgcagcactcgaagactctctgtgt 328
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 CTTCTCTGACTGATTCAGGATTATTGGCTTTGATCTCTGATCTCTTCCTTTTAA 401

Qy 329 actcgtggcagcatagctgttgcctgtgctgggacacctgtgtactac 383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 400 ACTGGTTATAGCAGAGGTACGGCGCTCCCTCTTCGCGAGGCGCTGTTCACAAC 346

```

```

RESULT 30
AW733988/c 544 bp mRNA EST 24-APR-2000
LOCUS
DEFINITION
sk79b06.y1 Gm-cl016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl016-10284 5' similar to TR:O22152 O22152 F4L23.30 PROTEIN. ;,
mRNA sequence.
ACCESSION AW733988
VERSION
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 544)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
TITLE Public Soybean EST Project
JOURNAL

```



**COMMENT**

Contact: Robert B. Weiss  
University of Utah  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112 USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0136 row: D column: 02  
Seq primer: CACACGAGAACGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 628.

## FEATURES

source

237

161 a	106 c	112 q	165 t
-------	-------	-------	-------

BASE COUNT  
ORIGIN

	Query Match	7.5%;	Score 33.4;	DB 242;	Length 628;	
	Best Local Similarity	52.5%;	Pred. No. 11.			
	Matches 73; Conservative	0;	Mismatches 56;	Indels 0;	Gaps 0;	
Oy	282	catttgagtgcgtttaccogtcgacgaactctccttgttactcgctggcacg	341			
DG	579	CAGTGATGTGGCGCTTGTATTCTGTCGTGGCAGGATGCCGCCTCTCTACCCGCTCTGCAGC	520			

Qy	342	atagtcgttttgcgcctggctctgggggacacactgtgtactacagctactcgtttcccttt	401
Db	519	TGTCCTCTCACAAACCATGTCTGTGCAAGACTTTCAGACACACGGTACTTGGCTCTCCCAT	460

QY	402
Dp	459

22  
23  
24

**LOCUS**  
**DEFINITION**

KEYWORDS  
SOURCE  
ORGANISM

homo sapiens  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

REFERENCE  
 1 (bases 1 to 933)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
 Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LLCM384 row: c column: 03  
 High quality sequence stop: 678.

#### FEATURES

source

1. 933  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3689402"  
 /clone\_lib="NIH\_MGC\_39"  
 /tissue\_type="adenocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: pancreas; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 148 a 316 c 319 g 150 t  
 ORIGIN

Query Match 7.5%; Score 33.4; DB 137; Length 933;  
 Best Local Similarity 58.6%; Pred. No. 12;  
 Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
 Qy 290 tgcgtttaccgtgcagcactcgaagacttcctgtgtactcgtgcagcatagtctg 349  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 422 TGCCTCAGACCCGGCTGGCGCTGCCATGCCGCTGCTGTTGGGCTGGCGCTGTGCTG 363  
 Qy 350 ttgtctacctgtctggggacacctgtgtactacagtac 388  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 362 TTAGTGCCTGGTCCCGCAGGAGCTGTCTGTGCAGTGC 324

RESULT 33  
 BE871340/c  
 LOCUS BE871340 945 bp mRNA EST 20-OCT-2000  
 DEFINITION 60149457F1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3853334 5', mRNA sequence.  
 ACCESSION BE871340  
 VERSION BE871340.1 GI:10320116  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 945)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
 Tissue Procurement: ATCC

CDNA Library Preparation: The I.M.A.G.E. Consortium, Inc.  
 DNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>  
 Plate: LHAM9577 row: a column: 15  
 High quality sequence stop: 656.

#### FEATURES

source

1. 945  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3853334"  
 /clone\_lib="NIH\_MGC\_65"  
 /tissue\_type="adenocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."

BASE COUNT 170 a 363 c 300 g 112 t  
 ORIGIN

Query Match 7.5%; Score 33.4; DB 141; Length 945;  
 Best Local Similarity 58.6%; Pred. No. 12;  
 Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 290 tgcgtttaccgtgcagcactcgaagacttcctgtgtactcgtgcagcatagtctg 349  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 392 TGCCTCAGACCCGGCTGGCGCTGCCATGCCGCTGCTGTTGGGCTGGCGCTGTGCTG 333  
 Qy 350 ttgtctacctgtctggggacacctgtgtactacagtac 388  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 332 TTAGTGCCTGGTCCCGCAGGAGCTGTCTGTGCAGTGC 294

RESULT 34  
 AV327005/c

LOCUS AV327005 225 bp mRNA EST 11-NOV-1999  
 DEFINITION AV327005 RIKEN full-length enriched, adult male medulla oblongata Mus musculus cDNA clone 6330416119 3', mRNA sequence.

ACCESSION AV327005  
 VERSION AV327005.1 GI:6367057  
 KEYWORDS EST.  
 SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 225)  
 AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Harada, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takanashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Konno, H., et al. 1999)  
 Unpublished (1999)  
 Contact: Yoshihide Hayashizaki

TITLE  
 JOURNAL  
 COMMENT

Genome Exploration Research Group, Life Science Tsukuba Center,  
 Genome Science Laboratory  
 The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: +81-298-36-9013

Fax: +81-298-36-9098

Email: [genome-res@rtr.riken.go.jp](mailto:genome-res@rtr.riken.go.jp)

URL: <http://genome.rtr.riken.go.jp/>

Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
 Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

This clone was previously sequenced on the 5' end only, this new data is from the 3' end  
High quality sequence stop: 286.

Location/Qualifiers  
1. .287

```

1. .207
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1382594"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"

```

Noti-Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I, and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. #

Query Match 7.5%; Score 33.2; DB 23; Length 287;  
Best Local Similarity 51.3%; Pred. NO. 10;

Query Match 7.5%; Score 33.2; DB 23; Length 287;  
Best Local Similarity 51.3%; Pred. NO. 10;

Qy	44	tctgaaggaaagacaagaacactccaagcgggacacgctactgtagtgcacgccga	103
Db	103	TTTGGGGGAAAAATAAAACCCCCCCCCCACCACCCCAAGGAATTTTGTGTGGGGGGG	162

QY      164    gtgaagccgtagtttgtacaagaactaaacaa    193  
             |     |     |     |     |     |     |     |  
Db      223    GGGACCCCTTTTAAAAAACCCTAA    252

RESULT	36				
AZ640083/c					
LOCUS	AZ640083	742 bp	DNA	GSS	14-DEC-2000
DEFINITION	1M0502K04F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0502K04 F, DNA sequence.				

ACCESSION  
VERSION  
KEYWORDS

SOURCE	house mouse.
ORGANISM	Mus musculus

ORGANISM MUS MUSCULUS  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Sciurognathi; Muridae; Murinae; Mus;

## REFERENCE AUTHORS

**AUTHORS**  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.,  
and Wright, D., Weiss, R.

**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
**JOURNAL** Unpublished (2000)

**JOURNAL**  
**COMMENT**

University of Utah Genome Center

University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

AUTHORS	JOURNAL
TITLE	COMMENT
...	...

## FEATURES

Db 546

D<sub>b</sub> 606

**ORGANIS**

JOURNAL

## FEATURES

[illegible]





Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

Sequencing the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589  
Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Fax: (202) 610-3667  
Email: [jwallace@u.washington.edu](mailto:jwallace@u.washington.edu)

## Sequence Tagged Connector

sequence tagged connector  
plate: 3212 row: P column: 10

Race: 5212 IO  
 Class: BAC endg

Class: BAC end  
High quality sequence stop: 437.

Location/Qualifiers

1. 437

```

T: 43 /
/organism="Homo sapiens"

```

```
/OligactisIII=HUMO sapient  
/db xref="taxon.9606"
```

```
/u01_xref=caxon:9606"
/cclone="Plate=3212 Col=10 Row=P"
```

/cclone\_lib="CTT Approved Human Genomic Screen Library D"

```

"save"major/
clone_lib

```

```
/sex="male"  
/note:"Organ: norm: Vector: polygenic: psc clones in
```

```
/note="Organ:
E-Coli DW10B"
```

E-Coli DH10B <sup>+</sup>		
72 a	144 c	124 q 96 t 1 others

## ORIGIN

```
Query Match      7.3%; Score 32.2; DB 225; Length 437;
Best Local Similarity 56.0%; Pred. No. 24;
Matches 61; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
```

Qy 302 tgcagcactcgaagacttccttgctgtactcgctggcagcatagctctgtttgctgcctgg 361

[illegible]

Db 241 TCCGGCGGAGCTGTGTTCTGTCGGCGTACTCACTGTCAATCGCTGCCGTGCCTCA

[illegible]

Search completed: September 7, 2001, 17:04:20  
Job time: 2444 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 7, 2001, 16:16:26 ; Search time 114.83 Seconds  
(without alignments)  
2427.835 Million cell updates/sec

Title: US-09-540-234-1  
Perfect score: 444  
Sequence: 1 gggtcacggcccttatgt.....tgtactcgtgtctgttcgag 444

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33.6	7.6	272	21 AAC70962	Single nucleotide
2	33.6	7.6	272	21 AAC70965	Single nucleotide
3	33.6	7.6	272	21 AAC70992	Single nucleotide
4	33.6	7.6	272	21 AAC71013	Single nucleotide
5	33.6	7.6	272	21 AAC71019	Single nucleotide
c 6	33.4	7.5	1894	21 AAZ50691	Human prostate der
c 7	33.4	7.5	1894	21 AAC83261	Gland-specific Ets
c 8	33.4	7.5	3317	22 AAC83266	Gland-specific Ets
c 9	32.8	7.4	1713	20 AAZ10357	Partial genomic se
10	31.6	7.1	2350	9 AAN82401	A and A' subunits
11	31.6	7.1	2350	12 AAQ14687	Factor XIII subuni

12	31.6	7.1	3425	21 AAC78122	Human cancer assoc
13	31.6	7.1	3905	8 AAN70461	Sequence encoding
14	31.6	7.1	3905	13 AAQ25893	Human Factor XIII
15	31.2	7.0	946	21 AAC41546	Arabidopsis thalia
16	31.2	7.0	1278	21 AAC41498	Arabidopsis thalia
17	31.2	7.0	1440	21 AAC49866	Arabidopsis thalia
c 18	29.4	6.6	1868	17 AAT39340	DNA sequence for m
c 19	29.4	6.6	21636	21 AAAS5966	Human G713 3'-end
c 20	29.2	6.6	1251	20 AAZ19436	M. tuberculosis an
c 21	29.2	6.6	1251	20 AAZ19224	M. tuberculosis re
c 22	29.2	6.6	3515	20 AAX03038	Human IL-1ra BAC c
c 23	29.2	6.6	11901	20 AAX02998	Human IL-1ra BAC c
24	29	6.5	411	21 AAC26317	Human secreted pro
25	29	6.5	488	21 AAC06849	Human secreted pro
26	29	6.5	676	21 AAC77597	Human ORFX ORF3152
27	29	6.5	1193	21 AAC52561	Arabidopsis thalia
28	29	6.5	1889	21 AAL13984	Taxus canadensis g
29	29	6.5	2812	21 AAC59139	Human secreted pro
c 30	28.8	6.5	1889	21 AAC36342	Arabidopsis thalia
c 31	28.8	6.5	10732	21 AAL10594	Gene encoding a su
c 32	28.6	6.4	647	22 AAF74196	DNA encoding eviro
c 33	28.6	6.4	5113	21 AAC83722	Rat laminin 5 cDNA
c 34	28.6	6.4	5264	21 AAC83721	Rat laminin 5 cDNA
c 35	28.4	6.4	2288	22 AAF97897	Human secreted pro
c 36	28	6.3	815	22 AAF22709	Human gastric canc
c 37	28	6.3	50000	21 AAAG6363	Polymorphic repeat
c 38	27.8	6.3	1253	20 AAX57500	Rat Ptx3 cDNA. Ra
c 39	27.8	6.3	1392	20 AAX59545	Nucleotide sequenc
c 40	27.8	6.3	1663	20 AAX13214	Enterococcus faeca
c 41	27.8	6.3	4740	10 AAN91159	DNA sequence downs
c 42	27.8	6.3	6666	12 AAQ10416	Mutant protease ge
c 43	27.8	6.3	6675	12 AAQ10414	Mutant protease ge
c 44	27.8	6.3	6675	12 AAQ10415	Mutant protease ge
c 45	27.8	6.3	6675	12 AAQ10411	Mutant protease ge

ALIGNMENTS

RESULT 1  
ID AAC70962 standard; DNA; 272 BP.  
XX AC AAC70962;  
XX DT 09-FEB-2001 (first entry)  
XX DE Single nucleotide polymorphism containing sequence #264.  
XX KW Single nucleotide polymorphism; SNP; human; genetic disease;  
KW disease susceptibility; cardiovascular system; endocrine system;  
KW neurological system; forensic testing; paternity testing; ds.  
XX OS Homo sapiens.  
XX PN WO200058519-A2.  
XX PD 05-OCT-2000.  
XX PF 30-MAR-2000; 2000WO-US08440.  
XX PR 31-MAR-1999; 99US-0127248.  
XX (WHEE) WHITEHEAD INST BIOMEDICAL RES.  
XX (AFFY-) AFFYMETRIX INC.  
XX PI Altshuler D, Cargill M, Daley GO, Ireland JS, Lander ES;  
PI Lipshutz RJ, Patil N, Sklar P;  
XX DR WPI; 2000-611722/58.  
XX PT Nucleic acid selected from one of 106 genes comprising single  
PT nucleotide polymorphisms, allele-specific oligonucleotides to the genes





	Query Match	7.6%	Score 33.6;	DB 21;	Length 272;
	Best Local Similarity	59.4%;	Pred. NO. 0.15;		
	Matches 57;	Conservative 0;	Mismatches 39;	Indels 0;	Gaps 0;
135	tgctctcaaacatgccccctgagcgtcggtggaagccgtagtgttcacaaactaacac	194			
108	tcctttaaagaacacctgcgaaatgctcgtggtacacctggatggctcctgggtaacaa	167			
195	gccaatggcgctgatgttatcctggagccattcgtgtg	230			
168	accaatggaagaagatgttcctgataaaccttgggttg	203			

RESULT	6
AAZ50691/c	
ID	AAZ50691 standard; cDNA; 1894 BP.
XX	
XX	AAZ50691;
AC	
AC	
XX	
DT	23-MAY-2000 (first entry)
XX	
XX	
DE	Human prostate derived Ets factor cDNA.

KW	Prostate Derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW	loss of heterozygosity; chromosomal translocation; linkage analysis;
KW	cystostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW	antibacterial; vulnery; neuroprotective; antiparkinsonian; nootropic;
KW	anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW	tumour marker; diagnosis; treatment; prostate cancer; blood coagulation;
KW	autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW	neoplasm; microbial infection; tissue regeneration; heart attack;
KW	scarring; food additive; preservative; ss.
XX	
OS	Homo sapiens.
OS	
XX	
XX	
FH	Key
FT	Location/Qualifiers
FT	416..1423
FT	CDS
FT	/tag= a
FT	/product= "Human prostate derived Ets factor"
FT	/note= "Expressed in prostate epithelium"
FT	

PS The present cDNA sequence encodes a human prostate derived Ets factor  
XX (PDEF), isolated from human prostate carcinoma cDNA library  
CC (ATCC No.203072). PDEF gene expression is highest in tissues involved in  
CC androgen sensitivity, with enriched expression in prostate epithelium  
CC cells. PDEF gene is mapped to the human chromosome 6p21.3  
CC region that is associated with loss of heterozygosity and chromosomal  
CC translocations in various human cancers. PDEF has cytostatic, cardiant,  
CC immunosuppressive, cerebroprotective, fungicide, antibacterial,

CC vulnery, neuroprotective, antiparkinsonian, nootropic, anabolic,  
CC antiinflammatory and anorectic activity. PDEF polynucleotides are useful  
CC in linkage analysis as markers, as hybridisation probes for differential  
CC identification of the tissues or cell types and as polymorphic markers  
CC for forensic purposes. PDEF is useful as prostate-specific tumour marker  
CC for the diagnosis and treatment of prostate cancer. PDEF sequences are  
CC useful for treating autoimmune disorders, haematopoietic, blood  
CC coagulation, immune and nervous system disorders, hyperproliferative  
CC disorders like, neoplasms and microbial infections, heart attacks,  
CC stroke, scarring and for tissue regeneration. They are also useful as  
CC food additives or preservatives.

Query Match	7.5%	Score 33.4;	DB 21;	Length 1894;
Best Local Similarity	58.6%;	Pred. No. 0.43;		
Matches 58;	Conservative 0;	Mismatches 41;	Indels 0;	Gaps 0;

  

QY	290	tgctgtttac	gtgcagc	ctgaag	acttctct	gtgtgt	tactcgt	ctgcgtgc	acgatag	ctgtg	349
Db	443	TGCTCAGAC	CCGGCTG	CGCGTGC	CCATG	CGCGTGT	TTGGGCT	GGCGCTG	GTGTCTG		384

  

QY	350	tttgtgtc	ctgtgt	ctgtgtg	gggacac	ctgtgt	tactac	agatc		388
Db	383	TTAGCTGC	CTTGGT	GCCAGG	AGTGTCT	GTGTGC	GTGAGT	GC		345

RESULT	7
AAC83261/c	
ID	AAC83261 standard; DNA; 1894 BP.
XX	
AC	AAC83261;
XX	
DT	16-MAR-2001 (first entry)
XX	
DE	Gland-specific Ets transcription factor (GSEF) nucleotide sequence.
XX	
KW	Transcription factor; gland-specific Ets transcription factor; GSEF;
KW	metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW	leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.

This invention relates to a method for the detection and determination of the metastatic potential of a cell. The method comprises detecting a gland-specific Ets transcription factor (GSEF) gene product in a test sample. Detection of a GSEF gene product in the test sample in amount lower than that in a normal cell, is indicative of a cell with high metastatic potential. The method is useful for determining the metastatic potential of a cell, for the diagnosis and prognosis of cancer as well as

sample. Detection of a GSEF gene product in the test sample in amount lower than that in a normal cell, is indicative of a cell with high metastatic potential. The method is useful for determining the metastatic potential of a cell, for the diagnosis and prognosis of cancer as well as grading and staging of cancers by detecting GSEF expression in a biological test sample. The method may also be used to monitor patients having a predisposition to develop a particular cancer. GSEF polypeptides are useful for producing antibodies, in cancer diagnosis, prognosis, grading, staging and management of breast and prostate tumours, and in detecting polymorphisms in the sequence. GSEF genes and proteins are also useful in gene therapy. GSEF gene product expression levels can be used in conjunction with any tissue in which an alteration in GSEF gene product expression levels is associated with development of a cancer-associated phenotype. Cancers, which can be monitored include adenocarcinomas, Wilm's tumour, retinoblastoma, melanomas, colorectal cancers of the prostate, cervix, lung and colon, sarcomas, myosarcomas, lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence represents the cDNA encoding GSEF.

	Oy	290	tgcgttttacccgctcagcactgaagaaccttccttgctgtactcgctggcagcatagtctg	349
	Db	443	TGCTCAGACCCGGGGCTGGCGCTGCCATGCCTGCTGTGTTGGGCTGGCGGCTGTGTCG	384
	Oy	350	tttgcgtcctggctctggggacacctgtgtactacagtac	388
	Db	383	TTAGCTGCCCTGGTGGCCGAGGAGCTGCTGCTCACTGC	345

AC	AAC83266;	
XX		
XX	16-MAR-2001 (first entry)	
DT		
XX		
DE	Gland-specific Ets transcription factor (GSEF)	cdna sequence.

OS	Homo sapiens.
XX	
PN	WO200070092-A1.

PR 14-MAY-1999; 99US-0134112.  
XX  
PA (CHIR ) CHIRON CORP.  
VV

**Detecting metastatic and potential metastatic cancerous cells, useful for diagnosing, prognosing, grading and staging of cancers by detecting specific Ets transcription factor gene product in a biological sample.**

CC This invention relates to a method for the detection and determination of  
CC the metastatic potential of a cell. The method comprises detecting a  
CC gland-specific Ets transcription factor (GSEF) gene product in a test

CC deleted in liver cancer cells. Detecting deletion of this gene, or  
 CC absence of the expressed protein, indicates increased susceptibility  
 CC to cancer, or presence of cancer (particularly hepatocellular carcinoma;  
 CC cancer of prostate, colon/rectum or breast, or adenocarcinoma). The  
 CC DLC-1 gene, or its cDNA, can be used in gene therapy to replace lost  
 CC gene function, specifically for treating cancer; and to generate  
 CC knockout transgenic animals (as in vivo models of carcinogenesis).  
 XX  
 SQ Sequence 1713 BP; 433 A; 485 C; 451 G; 326 T; 18 other;

Query Match 7.4%; Score 32.8; DB 20; Length 1713;  
 Best Local Similarity 53.0%; Pred. No. 0.65;  
 Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 223 ttctgtgctgtttactgttagcgttgatgacttctgttactcgtcgttgagcc 282  
 DB 697 TCCGGGTCCTGGTACCGGGCTGGGTGTGTGACCGGCTGCTGTCTGTGATTGGCTGC 638  
 QY 283 atttgagtgctgtttacgcgtgcagcactcgaagacttccttctgttactcgtcgcagca 342  
 DB 637 TGCTTGCTGCTGGTCTGGTGAAGCTTAGACACGCGCTCTTTCCGGTACCATGGGACGT 578  
 QY 343 tagtctgtttgc 354  
 DB 577 TGATGTGTTGC 566

RESULT 10  
 AAN82401  
 ID AAN82401 standard; DNA; 2350 BP.  
 XX  
 AC AAN82401;  
 XX  
 DT 26-NOV-1990 (first entry)  
 XX  
 DE A and A' subunits of human Factor XIII.  
 XX  
 KW A sub-unit; factor XIII; blood clot stabilisation; fibrin polymer;  
 KW crosslink; scleroderma; haemorrhage; ulcerative colitis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 CDS 91..2284  
 FT /\*tag= a  
 FT /product=mature a subunit  
 FT 202..2284  
 FT /\*tag= b  
 FT /product=mature a' subunit  
 XX  
 XX AU8778694-A.  
 XX  
 PD 31-MAR-1988.  
 XX  
 XX 21-SEP-1987; 87AU-0078694.  
 XX  
 PR 19-SEP-1986; 86US-0909512.  
 XX  
 XX (ZYMO-) ZYMOGENETICA INC.  
 XX  
 XX Davie E, Seale RL, Ichinose A, Holly JA, Parker GE;  
 XX WPI: 1988-140637/21.  
 DR P-PSDB; AAP82920.  
 XX  
 XX New DNA sequences encoding sub-units of factor 13 - and corresponding  
 PT expression vectors and transformed host cells.  
 XX  
 XX Disclosure; ; p; English.  
 XX  
 CC The carboxyl-terminal Met (nucleotides 2281-2283) is followed by  
 CC a stop codon (TGA), 1535 bp noncoding sequence, and a potential

CC polyadenylation or processing signal of AATAAA. The polyadenylation  
 CC sequence was located 14 nucleotides upstream from the poly(A) tail  
 CC of 10 nucleotides. The poly(A) tail was present only in a second  
 CC cDNA clone, designated lambdaHFXIIa3.82.  
 CC A difference in the nucleotide sequence for the a subunit was found  
 CC at three positions when comparison of the cDNA inserts was made in  
 CC regions where overlapping sequences were obtained. Nucleotides  
 CC 2038, 2041, and 2727 contained A, C, and T resp. in  
 CC lambdaHFXIIa3.77, while lambdaHFXIIa3.82 contained G, C, and A  
 CC in the same positions. These differences results in a change in two  
 CC amino acids (Ile 680 and Gln 681 to Val and Glu), and could  
 CC represent a polymorphism that contributes to the microheterogeneity  
 CC in the a subunit of factor XIII (P.G. Board and M. Coggan,  
 CC Hum.Genet. 59:135-136).  
 CC When cultured the host cells will produce the polypeptide which  
 CC can be assembled to factor 13, which stabilises blood clots by  
 CC crosslinking fibrin polymers. Factor 13 is useful therapeutically, e.g.  
 CC in cases of scleroderma, haemorrhage, ulcerative colitis etc., and  
 CC can be prep'd. in large amts. without risk of viral contamination.  
 CC See also AAN82402 and AAN82403.  
 XX  
 SQ Sequence 2350 BP; 645 A; 553 C; 620 G; 532 T; 0 other;

Query Match 7.1%; Score 31.6; DB 9; Length 2350;  
 Best Local Similarity 62.8%; Pred. No. 1.9;  
 Matches 49; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 135 tcgtctcaaacatgcccctcgagacgtcggtgagcgtagttgttacaagactaacac 194  
 DB 2052 tcctttaaagaacacctgcgaatgtctgtgacacctggtgctcgtgagtaacaag 2111  
 QY 195 gccaatggcgtcgatgtt 212  
 DB 2112 accaatgaagaagatgtt 2129

RESULT 11  
 AAQ14687  
 ID AAQ14687 standard; CDNA; 2350 BP.  
 XX  
 AC AAQ14687;  
 XX  
 DT 14-FEB-1992 (first entry)  
 XX  
 DE Factor XIII subunit a.  
 XX  
 KW Factor XIII; subunit; antibody; diagnosis; thrombosis; mutant; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 CDS 91..886  
 FT /\*tag= a  
 FT /label= factor\_XIII  
 FT /note= "mature a subunit"  
 FT 202..811  
 FT /\*tag= b  
 FT /note= "mature a' subunit"  
 XX  
 XX WO9116931-A.  
 XX  
 PD 14-NOV-1991.  
 XX  
 XX 09-MAY-1991; 91WO-US03212.  
 XX  
 XX 18-MAY-1990; 90US-0525556.  
 PR 10-MAY-1990; 90US-0521805.  
 XX  
 XX (ZYMO-) ZYMOGENETICS INC.  
 XX  
 XX Bishop PD;  
 XX



```

XX PD 16-SEP-1987.
XX PF
XX PF 06-MAR-1987; 87EP-0103222.
XX PR 26-JUN-1986; 86DE-3621371.
XX PR 12-MAR-1986; 86DE-3608280.
XX PA (BEHW ) BEHRINGWERKE AG.
XX PI Grundmann U, Amann E, Zettlmeissl G;
XX DR WPI; 1987-258275/37.
XX DR P-PSDB; AAP70293.
XX PT New DNA sequence coding for factor 13A and expressed proteins -
XX PT useful as diagnostic reagents and for producing antibodies
XX PS Claim 1; Table 3, pp16-20; 30pp; German.
XX CC Human placental cDNA gene bank was screened by hybridisation with two
XX CC synthetic oligonucleotides, corresponding to partial AA sequences of
XX CC factor XIIIa (AAN70460, AAN70465). AAN70461 gives the coding strand
XX CC sequence of clones lambda-gt10-11 and lambda-gt10-12.
XX SQ Sequence 3905 BP; 1175 A; 865 C; 898 G; 967 T; 0 other;

Query Match 7.1%; Score 31.6; DB 8; Length 3905;
Best Local Similarity 62.8%; Pred. No. 2.5;
Matches 49; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 135 tcgtctcaaacatgccctcgagacgtcggtgaagccgtagttgtacaagactaacac 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2049 tcctttaaagaacccctcgaaatgtctggtggtacacctggtgctctggagtaacaag 2108

Qy 195 gccaatggcgctcgatgtt 212
    ||||| ||| |||||
Db 2109 accaatgaagaagatgtt 2126

RESULT 14
AAC25893
ID AAC25893 standard; cDNA; 3905 BP.
XX AC AAC25893;
XX DT 08-JAN-1993 (first entry)
XX DE Human Factor XIII coding sequence.
XX KW Factor 13; Factor XIIIa; plasma-transglutaminase; fibrinolytic;
XX KW fibrin stabilising factor; blood clotting; coagulation; ss.
XX OS Synthetic.
XX FT Key Location/Qualifiers
XX FT CDS 85..2283
XX FT /*tag="a
XX FT /product=" Factor_XIII

XX EP494702-A.
XX PN
XX XX
XX PD 15-JUL-1992.
XX PF
XX PF 06-MAR-1987; 87EP-0105735.
XX PR 12-MAR-1986; 86DE-3608280.
XX PR 26-JUN-1986; 86DE-3621371.
XX PA (BEHW ) BEHRINGWERKE AG.
XX PI Amann E, Grundmann U, Zettlmeissl G;

```

```

XX WPI; 1992-235765/29.
XX DR P-PSDB; AAR25385.
XX PT Pure recombinant protein with factor XIII activity - useful in
XX PT diagnosis and treatment of factor XIII defects
XX PS Example 1; Page 16-20; 30pp; German.
XX CC Mature human placenta RNA was isolated and converted to ds cDNA.
XX CC A cDNA library was then prepared in Lambda phage gt10. The library
XX CC was screened with two probes (see AAQ25891 and AAQ32258) based on the
XX CC sequences of peptide fragments from Factor XIII. One clone
XX CC (lambda gt10-12) hybridised with both probes and was found to
XX CC contain a 1704 bp Factor XIIIa sequence with an internal EcoRI site.
XX CC The two EcoRI fragments generated from the insert sequence (540bp and
XX CC 1164bp, respectively) were labelled and used to screen the cDNA
XX CC library for clones extending at both ends compared to lambda gt10-12.
XX CC The full-length sequence coding for the inactive Factor XIII
XX CC polypeptide was constructed from the sequences of overlapping clones.
XX CC See also AAQ25328-9.
XX SQ Sequence 3905 BP; 1175 A; 871 C; 893 G; 966 T; 0 other;

Query Match 7.1%; Score 31.6; DB 13; Length 3905;
Best Local Similarity 62.8%; Pred. No. 2.5;
Matches 49; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 135 tcgtctcaaacatgccctcgagacgtcggtgaagccgtagttgtacaagactaacac 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2049 tcctttaaagaacccctcgaaatgtctggtggtacacctggtgctctggagtaacaag 2108

Qy 195 gccaatggcgctcgatgtt 212
    ||||| ||| |||||
Db 2109 accaatgaagaagatgtt 2126

RESULT 15
AAC41546
ID AAC41546 standard; DNA; 946 BP.
XX AC AAC41546;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 32273.
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway;
XX KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PF 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.

```



PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143342.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.

```
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 7.0%; Score 31.2; DB 21; Length 946;
Best Local Similarity 66.2%; Pred. No. 1.7;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Oy 205 tggatgttcgagaccattcgtgtgctgtttactgtgttagcgttgatgacttccttgc 264
    ||||| ||||| || || ||||| ||||| || |||||
Db 472 tggacttattcatggtcatttgcattgcgttcattgtttagctcttcactactcttga 531

Oy 265 tgtactcc 272
    ||| | |
Db 532 tgtgttac 539

RESULT 16
AAC41498
ID AAC41498 standard; DNA; 1278 BP.
XX
AC AAC41498;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 32103.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
```

```
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 7.0%; Score 31.2; DB 21; Length 1278;
Best Local Similarity 66.2%; Pred. No. 2;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 205 tcgatgttatcgagccattcgctgtgttactgttagcgttgacttccttgc 264
||||| ||||| | ||||| || ||||| ||||| ||||| |||||
Db 741 tcgatcttatcatggtcattgcattgcattgcgtttactgtttactcttcactctcttga 800

QY 265 tgtactcc 272
||| |||
Db 801 tgtgttac 808

RESULT 17
AAC49866
ID AAC49866 standard; DNA; 1440 BP.
XX
AC AAC49866;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 62732.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
```

PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139819.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149920.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 23-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.

```
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 7.0%; Score 31.2; DB 21; Length 1440;
Best Local Similarity 66.2%; Pred. No. 2.1;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 205 tcgatgtatcggagcattcgtgtgttactgttagcgttgatgacttcttgc 264
||||| ||||| | ||||| ||| ||||| ||||| ||||| |||||
Db 741 tcgatattatcattgcttgcattgcattgcttactgtttagcttcttactcttga 800

QY 265 tgtactcc 272
||| | |
Db 801 tgtgttac 808

RESULT 18
AAT39340
ID AAT39340 standard; DNA; 1868 BP.
XX
AC AAT39340;
XX
DT 22-JAN-1997 (first entry)
XX
DE DNA sequence for macrophage receptor with collagenous domain.
XX
KW Macrophage receptor-collagenous; MARCO; immune defence;
KW phagocytosis; gene therapy; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..159
FT /*tag= a
FT CDS 160..1716
FT /*tag= b
FT 3'UTR 1717..1868
FT polyA_signal 1813..1818
FT /*tag= d
XX
PN WO9626219-A1.
XX
PD 29-AUG-1996.
XX
PF 19-FEB-1996; 96WO-FI00091.
XX
PR 21-FEB-1995; 95US-0392367.
XX
PA (ELOM/) ELOMAA O.
PA (KANG/) KANGAS M.
PA (TRYG/) TRYGGVASON K.
XX
PI Elomaa O, Kangas M, Tryggvason K;
XX
WPI; 1996-402322/40.
DR P-PSDB; AAW03561.
XX
PT New macrophage receptor, MARCO, that binds bacteria but not yeast
and related DNA, anti-sense probes and antibodies, involved in
```

```
PT immune defence and phagocytosis
XX
PS Claim 1; Fig 2A-B; 62pp; English.
XX
CC A DNA clone (AAT39340), designated Maf-6, codes for a novel
macrophage receptor having a collagenous structure, named MARCO
(AAW03561), a membrane-bound trimeric protein which probably plays
a role in host defence. The gene was isolated from a mouse
macrophage cDNA library by screening with human type XIII collagen
DNA. An isolated clone was used to rescreen the library, yielding
Maf-6. The DNA or its fragments can be used to produce recombinant
MARCO, as a probe to detect the MARCO gene (or mutations in it that
cause disease), or to correct defects in the MARCO gene.
XX
SQ Sequence 1868 BP; 524 A; 433 C; 576 G; 335 T; 0 other;

Query Match 6.6%; Score 29.4; DB 17; Length 1868;
Best Local Similarity 51.1%; Pred. No. 9.5;
Matches 69; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 33 ctagtgtgggtcattcgaaggaaagacaaagacacctccaaagccggacaggtactgttag 92
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1206 caagggtgacacagggaattccaaggacagaaagcacaagaggaatcaggatccccag 1265

QY 93 ctggcagcgcataatgtaaaacagtccttgccttttggaggtcgtctcaaacatgcccc 152
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1266 tctttaggcagaaggagacactggaagccctggctggcaggtcccaaggagaacc 1325

QY 153 tcgagacgtcgttga 167
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1326 tggcagatcgtgtca 1340

RESULT 19
AAA55966/c
ID AAA55966 standard; DNA; 21636 BP.
XX
AC AAA55966;
XX
DT 05-SEP-2000 (first entry)
XX
DE Human G713 3'-end of intron 2, exon 3 and 3'-regulatory region.
XX
KW Human; chromosome 13; G713; chromosome 13q31-q33; schizophrenia;
KW biallelic marker; polymorphism; central nervous disease; detection;
KW neuroleptic; G713 gene expression inhibitor; genotyping; ds.
KW brain disorder; psychiatric disorder; bipolar disorder; ds.
XX
OS Homo sapiens.
XX
PN WO200022122-A2.
XX
PD 20-APR-2000.
XX
PF 12-OCT-1999; 99WO-IB01730.
XX
PR 13-OCT-1998; 98US-0103955.
PR 30-OCT-1998; 98US-0106457.
XX
PA (GEST ) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen D, Essioux L;
XX
WPI; 2000-317979/27.
XX
PT Novel polynucleotide of human G713 gene useful for diagnosis and
prophylactic treatment of brain, psychiatric disorders like
schizophrenia and bipolar disorders
XX
PS Claim 1; Page 224-230; 271pp; English.
XX
CC The present invention describes an isolated, purified or recombinant
```









```
Query Match 6.5%; Score 29; DB 21; Length 488;
Best Local Similarity 54.1%; Pred. No. 6.8;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 ggttcacggcccttatgtatgcagagcgcctagtgggtcatctgaagaaagacaa 60
Db 130 gagtcaaggacgtgtaggaagcctaaagtctggaagctatctcagaagaagcactg 189
QY 61 agacactccaagccgacacgcgtactgtagctgacgcgcataatgg 109
Db 190 aaaaactcaaaagctgtcactgcagcaacagcaggatggagataatgg 238

RESULT 26
AAC77597
ID AAC77597 standard; cDNA; 676 BP.
XX
AC AAC77597;
XX
DT 08-FEB-2001 (first entry)
XX
Human ORFX ORF3152 polynucleotide sequence SEQ ID NO:6303.
XX
DE
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX
XX P-PSDB; AAB43388.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 5488-5489; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
```

```
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy.
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 676 BP; 223 A; 130 C; 146 G; 177 T; 0 other;

Query Match 6.5%; Score 29; DB 21; Length 676;
Best Local Similarity 54.1%; Pred. No. 7.9;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 ggttcacggcccttatgtatgcagagcgcctagtgggtcatctgaagaaagacaa 60
Db 425 ggtccaaaggacctgtaggaaagcctaaagtctgtggaagctatctcagaagaagcactg 484
QY 61 agacactccaagccgacacgcgtactgtagctgacgcgcataatgg 109
Db 485 aaaaactcaaaagctgtcactgcagcaacagcaggatggagataatgg 533

RESULT 27
AAC52561
ID AAC52561 standard; DNA; 1193 BP.
XX
AC AAC52561;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 71747.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
```

PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 20-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 07-MAY-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.

CC	production, isolation and purification of larger amounts of GGGP
CC	synthase in plants. GGGP synthase is useful in obtaining expression or
CC	enhanced expression of GGGP and other diterpenes, such as pacitaxel,
CC	useful as anticancer drugs. Isolated nucleic acids encoding GGGP
CC	synthase or hybridising with GGGP synthase encoding nucleic acids are
CC	used for identifying genes encoding GGGP synthase from microorganisms
CC	such as <i>Taxomyces andreanae</i> and <i>Penicillium raistrickii</i> .
XX	
SQ	Sequence 1889 BP; 535 A; 300 C; 469 G; 585 T; 0 other;
Query Match	6.5%; Score 29; DB 21; Length 1889;
Best Local Similarity	50.4%; Pred. No. 13;
Matches	71; Conservative 0; Mismatches 70; Indels 0; Gaps
QY	278 gagccatttgagtcgtttaccgtagcagcactgcagcactgaagacctctctgtctactgcgtgg 337
Db	1401 gagccaaagagagcgtcatcctcttgatcagataaagcgtgcacctttgttgggtcttg 1460
QY	338 cagcatagtcgttttgcctgcttcctggcgggacacctgtgtactacagactctgtttcc 397
Db	1461 cagattacattgcattcaggcaaaactgagacaaagctgtaagctattttacatc 1520
QY	398 cttgttagtggtactcttct 418
Db	1521 atctgtttttttgacatct 1541
RESULT	29
AAC59139	
ID	AAC59139 standard; cDNA; 2812 BP.
XX	
AC	AAC59139;
XX	
DT	02-FEB-2001 (first entry)
XX	
DE	Human secreted protein coding sequence SEQ ID NO: 42.
XX	
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW	antiallergic; hepatotropic; antidiabetic; antinflammatory; antiulcer;
KW	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; human; secreted protein; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200055177-A2.
XX	
PD	21-SEP-2000.
XX	
PF	09-MAR-2000; 2000WO-US06058.
XX	
PR	12-MAR-1999; 99US-0124145.
PR	03-DEC-1999; 99US-0168654.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben SM, Komatsoulis G;
XX	
DR	WPI; 2000-638177/61.
DR	P-PSDB; AAB28043.
XX	
PT	Novel nucleic acids encoding 49 human secreted proteins useful for
PT	treating cancers, hyperproliferative disorders, inflammatory disorders,
PT	neurological disorders and cardiovascular disorders -
XX	
PS	Claim 1; Page 335-336; 389pp; English.
XX	
CC	The invention relates to the isolation of genes AAB59108-A59156 encoding
CC	49 human secreted proteins AAB28012-B28060. The genes can be used to
CC	generate fusion proteins by linking to the gene for the human
CC	immunoglobulin G Fc portion (SEQID1) for increasing the stability of
CC	the fusion-protein as compared to the human protein only. The genes and

CC proteins are useful for preventing, ameliorating or treating medical  
CC conditions, e.g. by protein or gene therapy. The genes are isolated  
CC from a range of human tissues disclosed in the specification. The  
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)  
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.

XX Sequence 2812 BP; 902 A; 482 C; 638 G; 777 T; 13 other;

Query Match 6.5%; Score 29; DB 21; Length 2812;  
Best Local Similarity 54.1%; Pred. No. 16;  
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
Qy 1 ggggtccaggcccttatgtatgcaggagcgccttagtggtgcctcgaaggaaaggacaa 60  
Db 425 gagtcaaaaggacctgtaggagcctaaagtctgtggaagctattctagaagaagcactg 484  
Qy 61 agacacctcccaagcggacgcgtactgtagctggtgcgcgcgcataatgg 109  
Db 485 aaaaactcaaaagctgtcactgcagcaacagcaggatggagataatgg 533

RESULT 30  
AAC36342/C  
ID AAC36342 standard; DNA; 1889 BP.  
XX AAC36342;  
XX  
XX 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 13448.  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
XX Arabidopsis thaliana.

XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126284.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 28-APR-1999; 99US-0130891.  
XX 28-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.





CC whereas prior art cell lines have been created that produce but do not  
CC secrete only one or two chain laminins.  
XX  
SQ Sequence 5113 BP; 1449 A; 1245 C; 1331 G; 1088 T; 0 other;

	Query Match	6.4%	Score 28.6;	DB 21;	Length 5113;
	Best Local Similarity	61.3%;	Pred. No. 29;		
	Matches 46;	Conservative 0;	Mismatches 29;	Indels 0;	Gaps
Qy	226	gtgtgtgtttactgtgtgtgcgcttgatgaacttcctgtctactccgcttggagccatt	285		
Db	1832	gtttgtgttagcctgttgagactgattgtttctgtctagccttttccgtgtcattc	1773		
Qy	286	tgagtgcctgtttacc	300		
Db	1772	ttgcccctgttttaac	1758		

RESULT 34  
AAC83721/C  
ID AAC83721 standard; cDNA; 5264 BP.  
XX  
XX  
AC AAC83721;  
XX  
XX 02-MAR-2001 (first entry)  
XX  
XX Rat laminin 5 cDNA, SEQ ID NO: 9.  
XX  
XX  
KW Call; laminin 5; vulvulary; antiulcer; antiinflammatory; antidiabetic;  
KW Cat adhesion promoter; wound healing; ulcers; burn; skin graft;  
KW periodontitis; gingivitis; Type I diabetes; angiogenesis regulation; ss  
XX  
XX Rattus norvegicus.  
OS

PN	WO2000066731-A2.
XX	
XX	
PD	09 - NOV - 2000.
XX	
XX	
PF	28 - APR - 2000; 2000WO-US11459.
XX	
PR	30 - APR - 1999; 99US-0131720.
PR	21 - AUG - 1999; 99US-0149738.
PR	24 - SEP - 1999; 99US-0155945.
XX	
XX	
PA	(BIOS -) BIOSTATUM INC.
XX	
PI	Boutaud A;
XX	
DR	WFI; 2000-687538/67.
DR	P-PSDE; AAB48460.

PT Laminin 5-expressing cells, used to accelerate wound healing associated  
PT with diabetic foot ulcers, venous ulcers, pressure sores, skin surgery,  
PT burns, acute wounds and skin grafts -  
XX  
PS Claim 4; Page 97-104; 232pp; English.

The present sequence encodes a laminin 5 chain polypeptide. Recombinant laminin 5-expressing cells are used to accelerate wound healing, especially diabetic foot ulcers, venous ulcers, pressure sores, skin surgery, burns, acute wounds, skin grafts, corneal ulcerations, gastro-intestinal ulcers, periodontitis, and gingivitis. They are also used to improve the biocompatibility of medical devices, and to promote cell adhesion to a surface. They can be used for the ex vivo treatment of Type I diabetes. Laminin can also be used to regulate angiogenesis. The cell line produces and secretes recombinant heterotrimeric laminin, whereas prior art cell lines have been created that produce but do not secrete only one or two chain laminins.

Sequence 5264 BP; 1487 A; 1277 C; 1385 G; 1115 T; 0 other;

Query Match 6.4%; Score 28.6; DB 21; Length 5264;  
Best Local Similarity 61.3%; Pred. No. 29;  
Matches 46; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy	226	gtgtgctgttactgtgtgtagcgtctgtagactctctgtctactccgcctctggagccaatt	285
Db	1983	gTTTCTGTAGGCTGTGTAGAGCTGGATTGATTCTGTCTGAGCCTTTTCCGTGTCAATC	1924
Qy	286	tgagtgtctgtttacc	300
Db	1923	TTGGCCTCGTTTAAC	1909

RESULT 35  
AAAF97897  
ID AAAF97897 standard; CDNA; 2288 BP.

AA  
AC AAF97897;

01-JUN-2001 (first entry)  
Human secreted protein cDNA, SEQ ID NO: 24.

Human; secreted protein; immunomodulatory; antisclerotic;  
 dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;  
 vascular; anti-angiogenic; ophthalmological; neuroprotectant;  
 nootropic; anticonvulsant; antialzheimers; antiparkinsonian;  
 antimicrobial; vulnery; vaccine; gene therapy; cancer;  
 protein coordinate data; infection; ss.

OS Homo sapiens.

AA WO200121658-A1.

29-MAR-2001

XX  
PF  
22-SEP-2000.XX  
PR 24-SEP-1999. 99JTC-0155709XX  
DA  
(HITMA - ) HITMAN GENOME SCT INC[illegible]

PI Lafleur DW, Moore PA, Olsen H

XX  
DR WPI; 2001-235311/24.

AA Nucleic acids encoding 32 human secreted polypeptides, useful for  
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
PT disease and diabetic retinopathy -  
PT

XX  
PS  
Claim 1: Page 729; 890pp: English.

The present sequence encodes one of 32 novel human secreted polypeptides. The nucleic acid molecules and polypeptides they encode may be used in the prevention, diagnosis and treatment of diseases such as immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus and human immunodeficiency virus (HIV) infections), hyperproliferative disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic disorders (e.g. corneal graft neovascularisation and diabetic retinopathy), neurological disorders (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease), infectious diseases and/or for promoting wound healing, regeneration and/or chemotaxis. The nucleic acid molecules may be used to produce the secreted polypeptides. They may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples. The polypeptides may be used as antigens in the production of antibodies and in assays to identify modulators of their expression and activity.

Sequence 2288 BP; 717 A; 477 C; 490 G; 599 T; 5 other; 22 A

```
Query Match          6.4%; Score 28.4; DB 22; Length 2288;
Best Local Similarity 55.28; Pred. No. 23;
Matches 53; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 226 gtgtgctgttactgttagcgcttgactctctgtgactcctgtgactcctgtgagccatt 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 ggcgcggnntccgcgcgcgaggtcttctgactgctgtgcccgcgcgcaggtgtagccatg 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 286 tgagtgtgtttaccgtgcagcactcgaagacttcc 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 cagcgcgcgattccgcgcgcgcctcccaagcgtccc 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 36
AAF22709/C
ID AAF22709 standard; cDNA; 815 BP.
XX AC AAF22709;
XX XX
DT 26-MAR-2001 (first entry)
XX DE Human gastric cancer associated antigen nucleotide sequence SEQ ID:288.
XX KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX KW cancer associated antigen; cytostatic; cancer vaccine; ss.
XX OS Homo sapiens.
XX XX
FN WO200073801-A2.
XX XX
PD 07-DEC-2000.
XX XX
PF 26-MAY-2000; 2000WO-US14749.
XX XX
PR 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX XX
PI Obata Y;
XX XX
DR WPI; 2001-025274/03.
XX XX
PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX XX
PS Claim 50; Page 368; 799pp; English.
XX XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX XX
SQ Sequence 815 BP; 282 A; 132 C; 123 G; 256 T; 22 other;
```

```
Query Match          6.3%; Score 28; DB 22; Length 815;
Best Local Similarity 48.6%; Pred. No. 19;
Matches 70; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 179 gtacaagactaacacgcgaatggcgtcgatgttatcgagccattcgtgctgtttac 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 642 GTATAGGCAAGCAAAAGTNTTAACATGGCAGCCATTCTGCTTNCNGGGCCCTGTCC 583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Two human bacterial artificial chromosome (BAC) clones that included and flanked the human CTLA-4 locus were cloned and sequenced. The sequence data was assembled into a contiguous sequence that is presented in AAA96363-68. AAA96363-64 comprise BAC clone 22700, and AAA96365-68

```
Qy 239 tgttagcgttgatgacttcttctgtactccgcttgagccatttgagtgtgttta 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 582 TGTTTAGTCTTCTAGTGGGTTAAATTTTGTCTACTTTTGCAGAGAACTTCAGCAAGCTAGA 523
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 299 ccgtgcagcactcgaagacttctc 322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 522 ACTGGAAGGACTTTAAATTTTTCAT 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 37
AAA96363
ID AAA96363 standard; DNA; 50000 BP.
XX AC AAA96363;
XX XX
DT 08-FEB-2001 (first entry)
XX DE Polymorphic repeat microsatellite sequences present in the CTLA4 locus.
XX KW Autoimmune disease; polymorphic microsatellite repeat; PMR; CD28 gene;
XX KW ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CGRL; lupus;
XX KW insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy;
XX KW Graves disease; autoimmune hypothyroidism; myasthenia gravis; thymoma;
XX KW thyroiditis; postpartum thyroiditis; rheumatoid arthritis;
XX KW Hashimoto's disease; coeliac disease; ss.
XX OS Homo sapiens.
XX XX
FN Key Location/Qualifiers
FT satellite 5722..5746
    /*tag= a
    /note= "sara41/42 microsatellite repeat"
FT satellite 6550..6597
    /*tag= b
    /note= "sara43/44 microsatellite repeat"
FT satellite 19911..19956
    /*tag= c
    /note= "PW210/211 microsatellite repeat"
FT satellite 23904..23957
    /*tag= d
    /note= "sara45/46 microsatellite repeat"
FT satellite 27689..27780
    /*tag= e
    /note= "sara17/18 microsatellite repeat"
FT satellite 30766..30801
    /*tag= f
    /note= "sara19/20 microsatellite repeat"
XX XX
PN WO200056856-A2.
XX XX
PD 28-SEP-2000.
XX XX
PF 24-MAR-2000; 2000WO-US07938.
XX XX
PR 25-MAR-1999; 99US-0126215.
XX XX
PA (GEMY ) GENETICS INST INC.
XX XX
PI Ling V, Wu P, Gray GS;
XX XX
DR WPI; 2000-628257/60.
XX XX
PT Determining predisposition of humans to develop autoimmune disease
PT involves detecting polymorphic microsatellite repeat sequence within
PT human costimulatory receptor gene locus -
XX PS Claim 2; Page 67-82; 160pp; English.
XX XX
CC Two human bacterial artificial chromosome (BAC) clones that included
CC and flanked the human CTLA-4 locus were cloned and sequenced. The
CC sequence data was assembled into a contiguous sequence that is presented
CC in AAA96363-68. AAA96363-64 comprise BAC clone 22700, and AAA96365-68
```



PT Gene with restricted expression in mesencephalic dopaminergic

XX





```
Query Match      6.3%; Score 27.8; DB 12; Length 6675;
Best Local Similarity 53.2%; Pred. No. 60;
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 75 cggacacggctactgtagctggcgcgcataatggaacacagtcctcccttttgagg 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4132 CGGTGTGCCCAACATTAACTGCTCTGTGTGTCAGCACTAACCGTCCCTGTAACTTGATGG 4073

Qy 135 tcgctcacaacatgccctcgagacgctcggtgaagccgtagtttgtacaag 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4072 TCGCTTCAGAGTCTGATCCGAGCCATTGTTGAATTTCAAAATCAGTAAAGG 4022

RESULT 44
AAQ10415/c
ID AAQ10415 standard; DNA; 6675 BP.
XX
AC AAQ10415;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease gene (N166D).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria; ss.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT CDS 376..6262
FT /*tag= a
FT /product= mutant protease
FT mat_peptide 938..6262
FT /*tag= b
FT sig_peptide 376..937
FT /*tag= c
XX
PN EP411715-A.
XX
PD 06-FEB-1991.
XX
PF 02-AUG-1990; 90EP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEZU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX WPI; 1991-038622/06.
XX P-PSDB; AAR10561.
XX
PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
CC The wild-type L.lactis SK11 protease gene sequence was determined
CC by the applicant (EP-307011).
CC The mutant protease N166D having new cleavage specificities is
CC obtained by carrying out single amino acid substitutions.
CC This mutant may then be used to prepare hybrid protease genes,
CC the fusion being between a type I and a type III protease gene of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
XX
SQ See also AAQ10411-17 and AAQ10870-71.
XX
Sequence 6675 BP; 2007 A; 1502 C; 1556 G; 1610 T; 0 other;
```

```
Query Match      6.3%; Score 27.8; DB 12; Length 6675;
Best Local Similarity 53.2%; Pred. No. 60;
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 75 cggacacggctactgtagctggcgcgcataatggaacacagtcctcccttttgagg 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4132 CGGTGTGCCCAACATTAACTGCTCTGTGTGTCAGCACTAACCGTCCCTGTAACTTGATGG 4073

Qy 135 tcgctcacaacatgccctcgagacgctcggtgaagccgtagtttgtacaag 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4072 TCGCTTCAGAGTCTGATCCGAGCCATTGTTGAATTTCAAAATCAGTAAAGG 4022

RESULT 45
AAQ10411/c
ID AAQ10411 standard; DNA; 6675 BP.
XX
AC AAQ10411;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease gene (A137G/K138D).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria; ss.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT CDS 376..6262
FT /*tag= a
FT /product= mutant protease
FT mat_peptide 938..6262
FT /*tag= b
FT sig_peptide 376..937
FT /*tag= c
XX
PN EP411715-A.
XX
PD 06-FEB-1991.
XX
PF 02-AUG-1990; 90EP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEZU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX WPI; 1991-038622/06.
XX P-PSDB; AAR10557.
XX
PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
CC The wild-type L.lactis SK11 protease gene sequence was determined
CC by the applicant (EP-307011).
CC The mutant protease A137G/K138D having new cleavage specificities is
CC obtained by replacing two amino acids (nine bps).
CC This mutant may then be used to prepare hybrid protease genes,
CC the fusion being between a type I and a type III protease gene of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
XX
SQ See also AAQ10411-17 and AAQ10870-71.
XX
```

XX

SQ Sequence 6675 BP; 2007 A; 1502 C; 1557 G; 1609 T; 0 other;

Query Match 6.3%; Score 27.8; DB 12; Length 6675;  
Best Local Similarity 53.2%; Pred. No. 60;  
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 75 cggacacggtactgtagctggcagcgcataatggaaaacagtcaccttttgaggg 134  
Db 4132 CGGTGTGGCCAACTTAACCTTCTTGGTGTGACACTAACCGTCCCTGTAACTTGTATGG 4073  
QY 135 tggctcacaacatgccccctcgagacgtcggtgaaagccgtagttgtacaag 185  
Db 4072 TCGCTTCAGAGGTCTGATCCGAGCCATTGTGAATTTCAAAATCAGTAAAGG 4022

Search completed: September 7, 2001, 16:41:35  
Job.time: 1509 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 7, 2001, 14:58:56 ; Search time 1472.14 Seconds  
(without alignments)  
4665.100 Million cell updates/sec

Title: US-09-540-234-1  
Perfect score: 444  
Sequence: 1 gggccacggcccttatgt.....tgtactcgtgtctgttcgag 444

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_ba1.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_htgo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_other.\*  
33: em\_htg\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_om.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_rod.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	41.6	9.4	173843	77	AC084405	AC084405 Oryza sat
2	37.2	8.4	166723	75	AC074278	AC074278 Homo sapi
c 3	36.8	8.3	164289	76	AC079395	AC079395 Homo sapi
4	36.8	8.3	193308	81	AL391703	AL391703 Homo sapi
5	35.6	8.0	166847	84	CNS06C7S	AL390801 Homo sapi
c 6	35.6	8.0	167068	84	CNS06C88	AL391516 Homo sapi
c 7	35.6	8.0	176506	84	CNS06C70	AL390335 Homo sapi
8	35.2	7.9	1042	2	BCHMLYSN	X84058 B.cereus ge





agccattcgtgctgttactgttagcgcttgatgaccttccttgtctgctacgcgt 275  
||||| |||| | |||| | |||| | |||| |  
AGCCCTTTCGTGTAATCGAAATCTTCTGGTGCAACACAGTTTCTGCTGCTTCTCCTAT 161901

COMMENT

On Oct 12, 2000 this sequence version replaced nt:9964991.



	BASE COUNT	56652 a	38556 c	38995 g	58904 t	201 others
misc_feature						
		/note="assembly fragment:01657 fragment_chain:1" 40577..181999				
		/note="assembly fragment:03665 fragment_chain:1" 182100..193308				
misc_feature						
		/note="assembly fragment:03205 fragment_chain:1" 56652 a 38556 c 38995 g 58904 t				
ORIGIN						

Query Match	8.3%	Score 36.8	DB 81	Length 193308
Best Local Similarity	48.1%	Pred. No. 2		
Matches 104	Conservative	0	Mismatches 112	Indels 0
Gaps				
QY 200	tggcgtcgatgtatcaggagccattcgtgtgcttttactctgtgacgcttgatgacttc	259		
Db 121135	TGTTGCTCATGATTTTTCACGAGTGTTTCTAGT			
QY 260	cttgcgtctactccgccttggagccatttgagtgctgtttaccgctcgagcactcgagactt	319		
Db 121195	CTTTGGTTAGGTATATTCTTAAAGCAATTTTATTTATTTATTTTGCAGCTGTTGTAAAGG	121254		
QY 320	ccctgcgtactcgtcggcagcatagctctgttactcctgctcctgctggggacacactgtgta	379		
Db 121255	GTTTGAGTTCCTTGACTTGATTCAGCTTGCTGTTGCTGTGTGTTATAGGAGACTACTGA	121314		
QY 380	ctacagctactcgtttccctctttagtgggtactact	415		
Db 121315	TTTGCGTACATTAATTTTGTGCTCGAAACGTTACT	121350		

RESULT	5
CNS06C7S	
LOCUS	DNA HTG 08-JAN-2001
DEFINITION	Homo sapiens chromosome 14 clone R-728G21, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION	AL390801
VERSION	AL390801.3 GI:12140328
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 166847) Genoscope. Direct Submission Submitted (08-JAN-2001) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr) Web : www.genoscope.cns.fr On Jan 12, 2001 this sequence version replaced gi:9716875.
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

**IMPORTANT:** This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continue. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage, etc. . . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the sp6 end.

```

from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-629f19 (AC=AL3391516)
Downstream BAC (overlapping the SP6 end) : R-129M6 -----
Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 6.54x in Q20 bases; sum-of-contigs

```

Assembly program: Phrap; version 2.0  
Quality coverage: 6.54x in Q20 bases; sum-of-contigs

.....

## Overall quality chart :

```

Range      : bases
0          : 212
1 - 9      : 7
10 - 19    : 54
20 - 29    : 134
30 - 39    : 428
40 - 49    : 2710
50 - 59    : 5039
60 - 69    : 6436
70 - 79    : 17421
80 - 89    : 48805
90 - 99    : 85601

```

Percentage of bases with a quality value >= 40 : 99 %.

- \* NOTE: This is a 'working draft' sequence.
- \* This sequence will be replaced
- \* by the finished sequence as soon as it is available and
- \* the accession number will be preserved.

## FEATURES

```

source
1. .166847
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="14"
   /clone="R-728621"
   /clone_lib="RPCI-11"

```

BASE COUNT 54079 a 28819 c 29091 g 54858 t

## ORIGIN

```

Query Match      8.0%; Score 35.6; DB 84; Length 166847;
Best Local Similarity 51.4%; Pred. No. 4.8;
Matches 112; Conservative 0; Mismatches 99; Indels 7; Gaps 1;

Qy 209 tggatcgagaccattgctgtgtttactgtgttagcgttgacttcttgcgtgta 268
Db 65623 TGCCCTAGGAGCCATTACGACACAGGAATCTATTAGCATTGGTGGTGGT 65682

Qy 269 ctccgctggagccattgagctgtttaccgtgcagcactcgaagacttcttgcgt 328
Db 65683 GTCAGCCTCTCCCTTCAGTCTTCAGCCCTCG-----TGCTGCACCATCTTCT 65735

Qy 329 actcgcctggagcactgtagctgttctgtcctggttggggacacctgtgtactacgtac 388
Db 65736 ATCCACCCCTCAGTGATTCTCTCAGAGGATCTGTTGGAAATGCTGGTACTTCATATG 65795

Qy 389 ttccgttccctttagtggtgactacttcttagccatg 426
Db 65796 TTGTTCTCTTTTGGGAGAGCTCTTCTTGGCCATG 65833

```

## RESULT 6

```

CNS06C88/c      167068 bp      DNA      HTG      08-DEC-2000
LOCUS      Homo sapiens chromosome 14 clone R-629F19, *** SEQUENCING IN
DEFINITION      PROGRESS ***, in ordered pieces.
ACCESSION      AL391516
VERSION      AL391516.3 GI:11611190
KEYWORDS      HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
               1 (bases 1 to 167068)
               Genoscope.
REFERENCE      Direct Submission
AUTHORS      Submitted (07-DEC-2000) Genoscope - Centre National de Sequencage :
TITLE      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
JOURNAL      - Web : www.genoscope.cns.fr)
COMMENT      On Dec 9, 2000 this sequence version replaced gi:10046660.
               ----- Genome Center
               Center: Genoscope / Centre National de Sequencage
               Center code: GS

```

Web site: <http://www.genoscope.cns.fr/>  
 Contact: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr)

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continue. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage, etc. . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the SP6 end.

Upstream BAC (overlapping the T7 end) : R-728621  
 Downstream BAC (overlapping the SP6 end) : R-757K12 (AC-AL390335)  
 ----- Summary Statistics

Assembly program: Phrap; version 2.0  
 Quality coverage: 6.72x in Q20 bases; sum-of-contigs

## Overall quality chart :

```

Range      : bases
0          :
1 - 9      : 12
10 - 19    : 30
20 - 29    : 60
30 - 39    : 217
40 - 49    : 1822
50 - 59    : 3286
60 - 69    : 6904
70 - 79    : 20683
80 - 89    : 55200
90 - 99    : 78854

```

Percentage of bases with a quality value >= 40 : 99 %.

- \* NOTE: This is a 'working draft' sequence.
- \* This sequence will be replaced
- \* by the finished sequence as soon as it is available and
- \* the accession number will be preserved.

## FEATURES

```

source
1. .167068
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="14"
   /clone="R-629F19"
   /clone_lib="RPCI-11"
   /note="matching EMBL:G30630
   RHD:RH37984
   dbSTS:STS29655
   Identified using the e-PCR software (G. Schuler)"

```

BASE COUNT 55505 a 29483 c 29391 g 52689 t

## ORIGIN

```

Query Match      8.0%; Score 35.6; DB 84; Length 167068;
Best Local Similarity 51.4%; Pred. No. 4.8;
Matches 112; Conservative 0; Mismatches 99; Indels 7; Gaps 1;

Qy 209 tggatcgagaccattgctgtgtttactgtgttagcgttgacttcttgcgtgta 268
Db 85600 TGCCCTAGGAGCCATTACGACACAGGAATCTATTAGCATTGGTGGTGGTGGT 85541

Qy 269 ctccgctggagccattgagctgtttaccgtgcagcactcgaagacttcttgcgt 328
Db 85540 GTCAGCCTCTCCCTTCAGTCTTCAGCCCTCG-----TGCTGCACCATCTTCT 85488

Qy 329 actcgcctggagcactgtagctgttctgtcctggttggggacacctgtgtactacgtac 388
Db 85487 ATCCACCCCTCAGTGATTCTCTCAGAGGATCTGTTGGAAATGCTGGTACTTCATATG 85428

Qy 389 ttccgttccctttagtggtgactacttcttagccatg 426
Db 85427 TTGTTCTCTTTTGGGAGAGCTCTTCTTGGCCATG 85390

```

7



**COMMENT**

3  
3  
4  
5  
6  
2

FEATURES SOURCE

```

1. :11500
/organism="Homo sapiens"
/db xref="taxon:9606"

```

Query Match	7.8%;	Score 34.8;	DB 75;	Length 180944;
Best Local Similarity	58.8%;	Pred. No. 8.5;		

repeated region

[illegible]

4  
3  
2  
1  
0  
9  
8  
7  
6  
5  
4  
3  
2  
1  
0

```
misc_feature 1044..1352
/note="match: STS: Em:G42844"
repeat_region 1044..1357
/note="AluSq repeat: matches 1..313 of consensus"
repeat_region 1360..1519
/note="FRAM repeat: matches 1..160 of consensus"
repeat_region 1586..1760
/note="L1M1 repeat: matches 5299..5484 of consensus"
misc_feature 1723..2467
/note="match: GSS: Em:AQ896083"
complement(2124..2394)
/note="match: GSS: Em:AQ313063"
2135..2387
/note="match: GSS: Em:AQ200628"
2136..2430
/note="match: GSS: Em:AQ351010"
2141..2370
/note="match: GSS: Em:AQ318962"
complement(2142..2532)
/note="match: GSS: Em:AQ314242"
complement(2143..2448)
/note="match: GSS: Em:AQ028497"
complement(2149..2519)
/note="match: GSS: Em:AF010903"
complement(2150..2517)
/note="match: GSS: Em:AQ341931"
complement(2155..2387)
/note="match: STS: Em:G29081"
complement(2155..2792)
/note="match: GSS: Em:AQ357800"
complement(2161..2428)
/note="match: GSS: Em:AQ764595"
complement(2161..2516)
/note="match: GSS: Em:AQ532045"
complement(2161..2470)
/note="match: GSS: Em:AQ169318"
complement(2161..2535)
/note="match: GSS: Em:B84560"
2162..2394
/note="match: GSS: Em:AQ413395"
2162..2517
/note="match: GSS: Em:AQ270015"
complement(2162..2456)
/note="match: GSS: Em:AQ060045"
2162..2456
/note="match: GSS: Em:B51613"
complement(2167..2387)
/note="match: GSS: Em:AQ281464"
complement(2167..2526)
/note="match: GSS: Em:AQ497805"
complement(2167..2524)
/note="match: GSS: Em:AQ497783"
complement(2167..2531)
/note="match: GSS: Em:AQ058261"
2167..2515
/note="match: GSS: Em:AQ269024"
complement(join(2168..2204,2642..2882))
/note="match: GSS: Em:AQ749070"
complement(2168..2504)
/note="match: GSS: Em:AQ535341"
complement(2278..2534)
/note="match: GSS: Em:AG012200"
2301..2510
/note="match: STS: Em:G10526"
complement(2306..2545)
/note="match: GSS: Em:AQ055853"
complement(2306..2535)
/note="match: GSS: Em:B91074"
complement(2306..2519)
/note="match: GSS: Em:AQ240974"
2314..2515
repeat_region
/note="AluSq repeat: matches 1..210 of consensus"
2314..2504
```

```
misc_feature
/note="match: STS: Em:HSC84B1"
complement(2318..2513)
/note="match: GSS: Em:AQ138787"
2318..2516
/note="match: GSS: Em:AQ013658"
2389..2517
/note="match: GSS: Em:AQ320309"
2391..2526
/note="match: STS: Em:G37384"
2391..2534
/note="match: GSS: Em:AQ741808"
2391..2514
/note="match: GSS: Em:AQ894888"
2391..2516
/note="match: GSS: Em:B58258"
2565..2892
/note="AluY repeat: matches 1..311 of consensus"
2642..2918
/note="match: GSS: Em:AQ788416"
2642..2885
/note="match: GSS: Em:AQ744233"
2893..3052
/note="FRAM repeat: matches 4..162 of consensus"
3061..3077
/note="weak data"
3085..3393
/note="AluY repeat: matches 1..311 of consensus"
3425..3638
/note="AluJo repeat: matches 3..204 of consensus"
join(<4098..4234,5355..5458,8081..8178,9319..9469,
9610..9818,10814..10911,11009..11094,12118..12199,
14212..14358,14804..15035,16421..17024)
/gene="CSEIL"
/note="match: cDNAs: Em:U33286 Em:AF053641 Em:AB036757"
/evidence=not_experimental
/product="dJ470L14.1.1 (isoform 1 of chromosome
segregation 1 (yeast homolog)-like (CSEIL) )"
4098..17024
/gene="CSEIL"
/gene="continues in clone dJ155G6 (Em:AL121903) as
dJ155G6.2.1
match: proteins: Sw:P55060 Tr:Q9PTU3 Tr:Q9XZU1"
/codon_start=1
/evidence=not_experimental
/product="dJ470L14.1.1 (isoform 1 of chromosome
segregation 1 (yeast homolog)-like (CSEIL) )"
/db_xref="CAC14081.1"
/db_xref="GI:10944131"
/translacion="VPKEHLVSIPLINHLQAESIVVHYAAHALERLFTMRGPNNA
TLFTAATIAPEVEILLNLKALFLPGSSENEYIMKAIMRSFSLQQAIIPIPTLIT
QLTQKLLAVSKNPKPHNFHFAICLSIRITCKANPAVAVFEEALFLVFTLQNL
DVQEFIPYFVQMSLLLETHKNDIPSSYMALFPHLLOPLWERNGNPALVRLQAEI
ERGSNTIASAAADKIGLLGVFKLIASKANDHOGVLLNSIIHEMPPEVDQYRKQI
FILLFORLQNSKTKFKISFLVINLCYIKYALALQAEIFDGIQPKFMGVLEKIIIP
Query Match 7.8%; Score 34.6; DB 89; Length 113836;
Best Local Similarity 58.1%; Pred. No. 9.7;
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
Qy 166 gaagcgtagttgtacaagactaacacgcaatggcgtcgatgttcgagcattc 225
|||||
Db 25988 GAGGTGGGAGTTTGTATATGGCAGCATAGCCATCTTCTTGATTTTGTAAACCTTC 25929
|||||
Qy 226 gtgtgctgttactgtgtagcgttgatgactctctgtact 270
|||||
Db 25928 CTGTGGTCTGAGTTCGTGGAACTTGTTACTTGTGTACTGTACT 25884
|||||
RESULT 12
```



```

AF123274/c
LOCUS       AF123274               397 bp    DNA                MAM                27-MAY-1999
DEFINITION  Physeter catodon strain macrocephalus microsatellite JB69 sequence.
ACCESSION   AF123274
VERSION     AF123274.1  GI:4894613
KEYWORDS    sperm whale.
SOURCE      Physeter catodon
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
            Physeteridae; Physeter.
REFERENCE   1 (bases 1 to 397)
            Bond, J.M. and Amos, B.
            Sperm whale microsatellite loci
            Unpublished
REFERENCE   2 (bases 1 to 397)
            Bond, J.M. and Amos, B.
            Direct Submission
            Submitted (22-JAN-1999) Molecular Ecology Group, Zoology, Cambridge
            University, Downing Street, Cambridge CB2 3ET, UK
FEATURES             Location/Qualifiers
     1..397
     /organism="Physeter catodon"
     /db_xref="taxon:9755"
     /sex="male"
     /tissue.type="skin"
     /dev.stage="immature"
     /note="Physeter macrocephalus
     primer A: tagatcttaggtaactgaac; primer B:
     cttcacctaggctactgtact; PCR conditions: 2 mins @94
     degrees x1, 1 min @94 degrees, 30 secs @47 degrees, 50
     secs @72 degrees x5, 45 secs @90 degrees, 1 min @51
     degrees, 1 min @72 degrees x25; reaction conditions: 1.5
     mM MgCl2, 60 mM TrisAC, 2.5% formamide"
     repeat_region   1..397
     /note="microsatellite JB69; 20 alleles identified in 274
     whales"
BASE COUNT      104 a      77 c      103 g      112 t      1 others
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
255
```

DEFINITION	Homo sapiens clone RP11-9D19, WORKING DRAFT SEQUENCE, 11 unordered pieces.
ACCESSION	AC021765
VERSION	AC021765.3 GI:10198414
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	1 (bases 1 to 152393)
JOURNAL	Homo sapiens, clone RP11-9D19
REFERENCE	Unpublished
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeAreltano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,I., Morrow,J., Naytor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Sep 20, 2000 this sequence version replaced gi:9123795. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> ----- Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a> Contact: <a href="mailto:sequence_submissions@genome.wi.mit.edu">sequence_submissions@genome.wi.mit.edu</a> ----- Project Information Center project name: L3002 Center clone name: 9.D.19 ----- Summary Statistics Sequencing vector: M13; M77815; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.960731 Consensus quality: 146933 bases at least Q40 Consensus quality: 149520 bases at least Q30 Consensus quality: 150601 bases at least Q20 Insert size: 157000; agarose-1p Insert size: 151393; sum-of-contigs Quality coverage: 4.6 in Q20 bases; agarose-1p Quality coverage: 4.7 in Q20 bases; sum-of-contigs ----- * NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence. * As soon as it is available and the accession number will be preserved. * 1 28910: contig of 28910 bp in length * 28911 29010: gap of 100 bp * 29011 31437: contig of 2427 bp in length * 31438 31537: gap of 100 bp * 31538 33952: contig of 2415 bp in length * 33953 34052: gap of 100 bp * 34053 38372: contig of 4320 bp in length

RESULT	15
AC011159/c	
LOCUS	
DEFINITION	
ACCESSION	







```
QY 308 actcgaagactcttctgtctactcgtcgtcgcagcatagctgttttctgctgctgtctggg 367
Db 35050 AGTAGGAGACACATTTGTTTAAAGCCATTAGACAGACTACTGGCAAGTACTGT 35109
QY 368 gacacctgtgtactacagctactcgttctcctttgttagtggtactacttctta 420
Db 35110 TACAAGTAATTAGCACAAACATTGGTTTACCGTATTATTGTTGTTATTATTA 35162

RESULT 20
AC073540/c DNA HTG 31-AUG-2000
LOCUS Homo sapiens chromosome 5 clone RP11-164A5, WORKING DRAFT SEQUENCE,
DEFINITION 16 unordered pieces.
ACCESSION AC073540
VERSION AC073540.2 GI:9954696
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189942)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189942)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 31, 2000 this sequence version replaced gi:8655975.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 469346
Center clone name: RPCI-11_164A5
-----
Summary Statistics
Consensus quality: 176984 bases at least Q40
Consensus quality: 184449 bases at least Q30
Consensus quality: 185674 bases at least Q20
Estimated insert size: 184750; agarose-fp estimation
Estimated insert size: 188442; sum-of-contigs estimation
Quality coverage: 6.05 in Q20 bases; agarose-fp estimation
Quality coverage: 5.93 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1317: contig of 1317 bp in length
* 1318 1417: gap of unknown length
* 1418 2526: contig of 1109 bp in length
* 2527 2626: gap of unknown length
* 2627 5034: contig of 2408 bp in length
* 5035 5134: gap of unknown length
* 5135 7507: contig of 2373 bp in length
* 7508 7607: gap of unknown length
* 7608 10242: contig of 2635 bp in length
* 10243 10342: gap of unknown length
* 10343 14775: contig of 4433 bp in length
* 14776 14875: gap of unknown length
* 14876 20332: contig of 5457 bp in length
* 20333 20432: gap of unknown length
* 20433 27996: contig of 7564 bp in length
* 27997 28096: gap of unknown length
* 28097 39155: contig of 11059 bp in length
```

```
* 39156 39255: gap of unknown length
* 39256 56163: contig of 16908 bp in length
* 56164 56263: gap of unknown length
* 56264 72833: contig of 16570 bp in length
* 72834 72933: gap of unknown length
* 72934 88357: contig of 15424 bp in length
* 88358 88458: gap of unknown length
* 88459 103315: contig of 14858 bp in length
* 103316 103416: gap of unknown length
* 103417 127722: contig of 24307 bp in length
* 127723 127822: gap of unknown length
* 127823 155816: contig of 27994 bp in length
* 155817 155916: gap of unknown length
* 155917 189942: contig of 34026 bp in length.
FEATURES
Location/Qualifiers
source 1..189942
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-164A5"
/clone_lib="RPCI human BAC library 11"
BASE COUNT 54100 a 40736 c 39831 g 53756 t 1519 others
ORIGIN
Query Match 7.6%; Score 33.8; DB 74; Length 189942;
Best Local Similarity 49.7%; Pred. No. 17;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 248 ctgatgactctctgtctactcgcgttgagccatttgagtgctgtttaccgtgcagc 307
Db 172334 CCTGCTGACACCTTAACCTTTAGTCCAGTAAGACCCATTTTGGAGTTTGGCCACCAAGAC 172275
QY 308 actcgaagactctctgtctactcgtcgcagcatagctgttttctgctgctgtctggg 367
Db 172274 AGTAGGAGACACATTTGCTGCTTTTAAAGCCATTTAGAACAGTACTTGGCAAGTACTGT 172215
QY 368 gacacctgtgtactacagctactcgttctcctttgttagtggtactacttctta 420
Db 172214 TACAAGTAATTAGCACAAACATTGGTTTACCGTATTATTGTTGTTATTATTA 172162

RESULT 21
AC026750/c
LOCUS Homo sapiens chromosome 5 clone RP11-469013, WORKING DRAFT
DEFINITION SEQUENCE, 21 unordered pieces.
ACCESSION AC026750
VERSION AC026750.3 GI:10305172
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 196716)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 196716)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Sep 26, 2000 this sequence version replaced gi:7459583.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 586810
Center clone name: RPCI-11_469013
-----
```

Summary Statistics					
Consensus quality: 181431 bases at least Q40					
Consensus quality: 190016 bases at least Q30					
Consensus quality: 191521 bases at least Q20					
Estimated insert size: 184750; agarose-fp estimation					
Estimated insert size: 194716; sum-of-contigs estimation					
Quality coverage: 3.35 in Q20 bases; agarose-fp estimation					
Quality coverage: 3.18 in Q20 bases; sum-of-contigs estimation					
* NOTE: This is a 'working draft' sequence. It currently					
* consists of 21 contigs. The true order of the pieces					
* is not known and their order in this sequence record is					
* arbitrary. Gaps between the contigs are represented as					
* runs of N, but the exact sizes of the gaps are unknown.					
* This record will be updated with the finished sequence					
* as soon as it is available and the accession number will					
* be preserved.					
* * *	1	1027:	contig of 1027 bp in length		
*	1028	1127:	gap of unknown length		
*	1128	2214:	contig of 1087 bp in length		
*	2215	2314:	gap of unknown length		
*	2315	3438:	contig of 1124 bp in length		
*	3439	3538:	gap of unknown length		
*	3539	5097:	contig of 1559 bp in length		
*	5098	5197:	gap of unknown length		
*	5198	6674:	contig of 1477 bp in length		
*	6675	6774:	gap of unknown length		
*	6775	8091:	contig of 1317 bp in length		
*	8092	8191:	gap of unknown length		
*	8192	9300:	contig of 1109 bp in length		
*	9301	9400:	gap of unknown length		
*	9401	11808:	contig of 2408 bp in length		
*	11809	11908:	gap of unknown length		
*	11909	14543:	contig of 2635 bp in length		
*	14544	14643:	gap of unknown length		
*	14644	17016:	contig of 2373 bp in length		
*	17017	17116:	gap of unknown length		
*	17117	21549:	contig of 4433 bp in length		
*	21550	21649:	gap of unknown length		
*	21650	27106:	contig of 5457 bp in length		
*	27107	27206:	gap of unknown length		
*	27207	34770:	contig of 7564 bp in length		
*	34771	34870:	gap of unknown length		
*	34871	45929:	contig of 11059 bp in length		
*	45930	46029:	gap of unknown length		
*	46030	62937:	contig of 16908 bp in length		
*	62938	63037:	gap of unknown length		
*	63038	79607:	contig of 16570 bp in length		
*	79608	79707:	gap of unknown length		
*	79708	95131:	contig of 15424 bp in length		
*	95132	95231:	gap of unknown length		
*	95232	110089:	contig of 14858 bp in length		
*	110090	110189:	gap of unknown length		
*	110190	134496:	contig of 24307 bp in length		
*	134497	134596:	gap of unknown length		
*	134597	162590:	contig of 27994 bp in length		
*	162591	162690:	gap of unknown length		
*	162691	196716:	contig of 34026 bp in length.		
FEATURES					
Location/Qualifiers					
1..196716					
source					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/chromosome="5"					
/clone="RP11-469013"					
/clone_lib="RPCI human BAC library 11"					
BASE COUNT	54117	a	40760	c	39860 g 53767 .t 8212 others
ORIGIN					
Query Match	7.6%		Score 33.8;	DB 70;	Length 196716;
Best Local Similarity	49.7%;		Pred. No. 17;		
Matches 86;	Conservative	0;	Mismatches	87;	Indels 0; Gaps
Qy	248	cttgatgactcctgtcgtactccgcctgttagagcatttgagtgtgtttaccqtcacg	307		



**ORGANISM**  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 273)  
**REFERENCE**  
 Ichinose, A. and Davie, E.W.  
**AUTHORS**  
 Characterization of the gene for the subunit of human factor XIII  
 (plasma transglutaminase), a blood coagulation factor  
**TITLE**  
 Proc. Natl. Acad. Sci. U.S.A. 85 (16), 5829-5833 (1988)  
**JOURNAL**  
 88320337  
**MEDLINE**



all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-525021 is from the library RPci-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

## FEATURES

```

    source            Location/Qualifiers
        1..163960
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone="RP11-525021"
            /clone_lib="RPci-11.2"
            3535..3669
                /note="3 copies 45 mer 85% conserved"
                complement(3583..3911)
                    /note="match: GSS: Em:AQ0705901"
                    3630..3907
                        /note="match: GSS: Em:AQ253322"
                        3664..3777
                            /note="match: GSS: Em:AQ317328"
                            3676..3843
                                /note="match: GSS: Em:B50322"
                                complement(3684..3839)
                                    /note="match: GSS: Em:AQ330998"
                                    3685..3926
                                        /note="match: GSS: Em:AQ526386"
                                        3730..3759
                                            /note="15 copies 2 mer at 96% conserved"
                                            3732..3759
                                                /note="7 copies 4 mer atat 100% conserved"
                                                3812..3839
                                                    /note="14 copies 2 mer at 96% conserved"
                                                    3814..3837
                                                        /note="6 copies 4 mer atat 100% conserved"
                                                        4094..4408
                                                            /note="7 copies 45 mer 74% conserved"
                                                            complement(4428..4749)
                                                                /note="match: STS: Em:HS451L15S"
                                                                4491..4580
                                                                    /note="2 copies 45 mer 87% conserved"
                                                                    4551..4614
                                                                        /note="32 copies 2 mer at 71% conserved"
                                                                        4565..5092
                                                                            /note="match: GSS: Em:AQ371839"
                                                                            4730..5241
                                                                                /note="12 repeat: matches 1749..2282 of consensus"
                                                                                5435..5547
                                                                                    /note="12 repeat: matches 2590..2749 of consensus"
                                                                                    5601..5680
                                                                                        /note="12 repeat: matches 1299..1380 of consensus"
                                                                                        6076..6362
                                                                                            /note="AluY repeat: matches 1..285 of consensus"
                                                                                            6802..7103
                                                                                                /note="11 repeat: matches 4734..5047 of consensus"
                                                                                                7329..7446
                                                                                                    /note="59 copies 2 mer tt 66% conserved"
                                                                                                    7520..8075
                                                                                                        /note="L1MC5 repeat: matches 7354..7935 of consensus"
                                                                                                        11177..11578
                                                                                                            /note="match: GSS: Em:B71797"
                                                                                                            11585..11651
                                                                                                                /note="MER5A repeat: matches 9..75 of consensus"
                                                                                                                11606..11887
                                                                                                                    /note="match: STS: Em:G17084"
                                                                                                                    11655..11738
                                                                                                                        /note="AluS repeat: matches 1..84 of consensus"
                                                                                                                        12974..13172
                                                                                                                            /note="12 repeat: matches 2214..2419 of consensus"
                                                                                                                            13344..13419
                                                                                                                                /note="19 copies 4 mer aaga 65% conserved"
                                                                                                                                13560..13941

```

```

repeat_region      /note="match: GSS: Em:AQ005912"
13994..14175
/note="MER5A repeat: matches 3..178 of consensus"
17440..17721
/note="L1MC5 repeat: matches 7628..7920 of consensus"
17831..18376
/note="L1PB1 repeat: matches 5599..6155 of consensus"
18377..18677
/note="AluSc repeat: matches 1..295 of consensus"
18678..19023
/note="L1PB1 repeat: matches 5257..5599 of consensus"
19191..19240
/note="L1M4C repeat: matches 1685..1737 of consensus"
19241..19540
/note="AluY repeat: matches 1..300 of consensus"
19541..19686
/note="L1M4C repeat: matches 1555..1685 of consensus"
21482..21897
/note="L1PA10 repeat: matches 5743..6161 of consensus"
22715..22790
/note="L1PA6 repeat: matches 6066..6143 of consensus"
22791..25265
/note="HERVK_9 repeat: matches 1625..4106 of consensus"
25267..25933
/note="HERVK_9 repeat: matches 1..661 of consensus"
25934..26423
/note="MER9 repeat: matches 1..511 of consensus"
26460..26775
/note="AluYa5 repeat: matches 1..311 of consensus"
26776..27304
/note="L1P repeat: matches 3090..3619 of consensus"
27305..27615
/note="AluY repeat: matches 1..309 of consensus"
27616..30704
/note="L1P repeat: matches 11..3090 of consensus"
30924..31188
/note="AluY repeat: matches 9..291 of consensus"
31832..32363
/note="MER7A4 repeat: matches 20..558 of consensus"
32365..32535
/note="L1M4 repeat: matches 5942..6118 of consensus"
32632..32776
/note="MER47A repeat: matches 5..150 of consensus"
32778..33125
/note="L1MEC repeat: matches 2404..2399 of consensus"
33187..33689
/note="L1M4 repeat: matches 2645..3148 of consensus"
34054..34446
/note="L1 repeat: matches 3547..3949 of consensus"
34511..37154
/note="L1M2 repeat: matches 3652..6308 of consensus"
37353..37603
/note="L1M4 repeat: matches 4795..5055 of consensus"
38056..38250
/note="MIR repeat: matches 50..227 of consensus"
38350..38377
/note="14 copies 2 mer ac 96% conserved"
38352..38375
/note="6 copies 4 mer acac 100% conserved"
38759..39104
/note="L2 repeat: matches 1770..2112 of consensus"
39449..39487
/note="MLT1J repeat: matches 111..149 of consensus"
40297..40706
/note="match: GSS: Em:AQ176196"
40494..40895
/note="MLT1A1 repeat: matches 1..365 of consensus"
41695..42062
/note="MLT1A2 repeat: matches 1..370 of consensus"
43042..43469
/note="L1PB2 repeat: matches 5711..6155 of consensus"
43605..43763
/note="MER5A repeat: matches 12..189 of consensus"

```

```

repeat_region 44885..44994
/note="L2 repeat: matches 2593..2690 of consensus"
repeat_region 44997..45420
/note="MLP2FB repeat: matches 1..414 of consensus"
repeat_region 46331..46356
/note="13 copies 2 mer tg 92% conserved"
repeat_region 46597..46636
/note="20 copies 2 mer ca 82% conserved"
repeat_region 46612..46635
/note="6 copies 4 mer acac 100% conserved"
repeat_region 46711..46754

Query Match 7.6%; Score 33.6; DB 90; Length 163960;
Best Local Similarity 59.4%; Pred. No. 20;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 135 tcgtctcaaacatgccctcgagcgtggtgaagcgcgtattgttgcacagactaacac 194
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109722 TCCTTTAAAGAAACCCCTGCCAAATGCTGGGTACACCTGGATGCTCTGGAGTAACAAG 109663
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 195 gccaatgcgtcgatgttatcgagcattcgtgtg 230
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109662 ACCAATGAAGAAGATGTCCCGTAACCTTGGGTG 109627

RESULT 26
AC025972 166402 bp DNA HTG 12-JAN-2001
LOCUS Homo sapiens chromosome 11 clone RP11-6K5 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 3 unordered pieces.
ACCESSION AC025972
VERSION AC025972.4 GI:12084073
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166402)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castler,A., Choepel,Y., Colangelo,W., Collins,S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (18-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 12, 2001 this sequence version replaced gi:8076808.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

```

```

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6107
Center clone name: 6_K5
----- Summary Statistics
Sequencing vector: M13; M7815; 40% of reads
Sequencing vector: Plasmid; n/a; 60% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165502 bases at least Q40
Consensus quality: 165912 bases at least Q30
Consensus quality: 166038 bases at least Q20
Insert size: 172000; agarose-fp
Insert size: 166202; sum-of-contigs
Quality coverage: 10.5 in Q20 bases; agarose-fp
Quality coverage: 10.9 in Q20 ba.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 137540: contig of 137540 bp in length
* 137541 137640: gap of 100 bp
* 137641 137742: contig of 102 bp in length
* 137743 137842: gap of 100 bp
* 137843 166402: contig of 28560 bp in length.
FEATURES
Location/Qualifiers
1..166402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-6K5"
/clone_lib="RP11-11 Human Male BAC"
1..137540
/note="assembly_fragment"
clone_end:SP6
vector_side:left
137641..137742
/note="assembly_fragment"
137843..166402
/note="assembly_fragment"
clone_end:r7
vector_side:right
BASE COUNT 49603 a 32681 c 33231 g 50687 t 200 others
ORIGIN

Query Match 7.6%; Score 33.6; DB 59; Length 166402;
Best Local Similarity 59.4%; Pred. No. 20;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 34 tagtgggtccatcgaagaaagcacaagcacctccaagccgacgcgtactgttagc 93
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77827 TAGTGGGGAACACTGAGTATAGGCATATTACAAACAACATGACATGCTATGTTAGTA 77886

Qy 94 tggcagcgcatatggaaacacagtcctgcctcttt 129
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77887 TTGAGTCCCATAGTGGACACGAGTCTTCAGCCTTGT 77922

RESULT 27
AC068860
LOCUS Homo sapiens chromosome 11 clone RP11-371120 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 6 unordered pieces.
ACCESSION AC068860
VERSION AC068860.3 GI:13357505
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```





factor, interacts with the androgen receptor and activates prostate-specific antigen gene expression  
J. Biol. Chem. 275 (2), 1216-1225 (2000)

## JOURNAL

## MEDLINE

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

factor, interacts with the androgen receptor and activates prostate-specific antigen gene expression  
J. Biol. Chem. 275 (2), 1216-1225 (2000)  
2 (bases 1 to 1894)  
Libermann, T.A., Oettingen, P., Kunsch, C. and Finger, E.  
Direct Submission  
Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical Center, 330 Brookline Ave., Boston, MA 02215, USA  
Location/Qualifiers  
1. 1894  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21.3"  
/tissue\_type="prostate"  
1. 1894  
/gene="PDEF"  
416. 1423  
/gene="PDEF"  
/note="prostate epithelium-specific"  
/codon\_start=1  
/product="Bts transcription factor PDEF"  
/protein\_id="AAC95296.1"  
/db\_xref="GI:4007418"  
/translation="MGSAASGLSSVSPSHLLPDDTVSRTGLEKAAAGAVLERDWS  
PSPATPQGLSAFYLSYFDMLYPEDSSAAKAPGASREPPPEQCPVIDSOAPA  
GSLDLPGLGLEHSLQVQSVVGVKDIETACKLNTADPMDWSPNVOKWLL  
WTEHQYRPLPPMKAFQELAGKELCAMSEEQFORSPGLGDLVHAHLDIWKAAMWKEK  
TSPGAIHYCASTSESWTDSEVSSCGQPHLWQFLKELLKPHSYGRFIRWLNKEK  
GIFKIEDSAQVRLWGIKRNPMNYDKLSRSIRQYKKGIIRKPDISQRLVYQFVHP  
I"

BASE COUNT 368 a 653 c 571 g 302 t

ORIGIN

Query Match 7.5%; Score 33.4; DB 88; Length 1894;

Best Local Similarity 58.6%; Pred. No. 21;

Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 290 tgcgtttaccgtgcagcactgaagacttcctgtgtactcgtgcagcagcatagctg 349

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 443 TGCTCAGACCCGGCGTGGCGCTGCCCATGCCGTGTTGGCGTGGCGGTGTCTG 384

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 350 ttgtgcctgctgtgggacacctgtgtactacagtac 388

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 383 TTAGCTGCTGTGGTCCAGGAGCTGCTGTCAGTGC 345

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 31

AB031549/c

LOCUS

AB031549 Homo sapiens PSE mRNA for prostate ets, complete cds.

DEFINITION

AB031549

ACCESSION

AB031549.1 GI:6721497

VERSION

AB031549.1

KEYWORDS

prostate ets.

SOURCE

Homo sapiens adult male prostate epithelium prostate cancer

cell\_line:PC-3 CDNA to mRNA.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (sites)

Yamada, N., Tamai, Y., Miyamoto, H. and Nozaki, M.

Cloning and expression of the mouse Pse gene encoding a novel Ets

family member

Gene 241 (2), 267-274 (2000)

JOURNAL

REFERENCE

2 (bases 1 to 1898)

Nozaki, M.

Direct Submission

Submitted (25-AUG-1999) to the DDBJ/EMBL/GenBank databases. Masami

Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,

Yamadaoka, Suita, Osaka 565-0871, Japan

(E-mail: mnozaki@biken.osaka-u.ac.jp, Tel: 81-6-879-8338,  
Fax: 81-6-879-8339)

Location/Qualifiers

1. 1898

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_line="PC-3"

/cell\_type="prostate cancer"

/chromosome="6"

/dev\_stage="adult"

/map="6p21.2-21.3"

/sex="male"

/tissue\_type="prostate epithelium"

/note="vector:pUC18"

406. 1413

/gene="PSE"

406. 1413

/gene="PSE"

/codon\_start=1

/product="prostate ets"

/protein\_id="BAA89543.1"

/db\_xref="GI:6721498"

/translation="MGSAASGLSSVSPSHLLPDDTVSRTGLEKAAAGAVLERDWS  
PSPATPQGLSAFYLSYFDMLYPEDSSAAKAPGASREPPPEQCPVIDSOAPA  
GSLDLPGLGLEHSLQVQSVVGVKDIETACKLNTADPMDWSPNVOKWLL  
WTEHQYRPLPPMKAFQELAGKELCAMSEEQFORSPGLGDLVHAHLDIWKAAMWKEK  
TSPGAIHYCASTSESWTDSEVSSCGQPHLWQFLKELLKPHSYGRFIRWLNKEK  
GIFKIEDSAQVRLWGIKRNPMNYDKLSRSIRQYKKGIIRKPDISQRLVYQFVHP  
I"

BASE COUNT 380 a 651 c 567 g 300 t

ORIGIN

Query Match 7.5%; Score 33.4; DB 85; Length 1898;

Best Local Similarity 58.6%; Pred. No. 21;

Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 290 tgcgtttaccgtgcagcactgaagacttcctgtgtactcgtgcagcagcatagctg 349

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 433 TGCTCAGACCCGGCGTGGCGCTGCCCATGCCGTGTTGGCGTGGCGGTGTCTG 374

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 350 ttgtgcctgctgtgggacacctgtgtactacagtac 388

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 373 TTAGCTGCTGTGGTCCAGGAGCTGCTGTCAGTGC 335

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 32

AX048159/c

LOCUS

AX048159 Sequence 12 from Patent WO0070092.

DEFINITION

AX048159

ACCESSION

AX048159.1 GI:11876975

VERSION

AX048159.1

KEYWORDS

human.

SOURCE

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 3317)

Kaufmann, J. and Xin, H.

Expression of ets-domain proteins in cancer

Patent: WO 0070092-A 12 23-NOV-2000;

Chiron Corporation (US)

Location/Qualifiers

1. 3317

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/note="GSEF-encoding sequence with promoter (Figs. 2A-2B)"

promoter

1. 1380

/note="Promoter"

1815. 2822

/note="GSEF coding sequence"

/codon\_start=1

/protein\_id="CAC19004.1"

CDS

```
/db_xref="GI:11876976"
/translation="MGSALFSLVSPSHLLPPDVTVRSGLEKAAAGVGLRRDWS
PSPATPEQGLSAFYLSYFDMLYPEDSWAAKAGASSREPEPEPCQVIDSQAPA
GSLDLYPGGTLLEHSLQVQSMVGEVLKDIETACKLNLITADPMDWSPSNVQKWL
WTEHQRLPMPGAFQELAGKCAESEEQRORSPLGGDVLHLDHDKSAAMWKEK
TSPGAIHLYCATSEESWTDEVSQSGQPIHLWFLKELLKPHSYGRFIRWLNEK
GIPIEDSAQVARLWGIKRNRPAMNYDKLSRSIRQYYKGIIRKPDISQRLVYQVHP
I"
BASE COUNT      710 a 1026 c  970 g  611 t
ORIGIN

Query Match      7.58; Score 33.4; DB 9; Length 3317;
Best Local Similarity 58.6%; Pred. No. 21;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 290 tgcgtttaccgtcagcactcgaagactctcttgcgttactcgtcgtcgcagcatagctg 349
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1842 TGCTCAGACCGCGCTGGCGTGCCTATCGCGCTGCTGTTGGCGTGGCGCTGCTG 1783
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 350 ttctcgtcgtgctggggacacactgtgtactacagtac 388
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1782 TTACTGCTGCTGCCCGCAGGAGCTGCTGCTGCAGTGC 1744
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 33
AC068504      66534 bp      DNA      HTG      03-MAY-2000
LOCUS      Homo sapiens chromosome 4 clone RP11-207J21 map 4, LOW-PASS
DEFINITION      SEQUENCE SAMPLING.
AC068504
VERSION      AC068504.1 GI:7684418
KEYWORDS      HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 4, clone RP11-207J21
Unpublished
2 (bases 1 to 66534)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zalnoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
```

```
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L10123
Center Clone name: 207_J_21
-----
* NOTE: This record contains 75 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 778: contig of 778 bp in length
* 779 878: gap of 100 bp
* 879 1660: contig of 782 bp in length
* 1661 1760: gap of 100 bp
* 1761 2542: contig of 782 bp in length
* 2543 2642: gap of 100 bp
* 2643 3416: contig of 774 bp in length
* 3417 3516: gap of 100 bp
* 3517 4294: contig of 778 bp in length
* 4295 4394: gap of 100 bp
* 4395 5194: contig of 800 bp in length
* 5195 5294: gap of 100 bp
* 5295 6093: contig of 799 bp in length
* 6094 6193: gap of 100 bp
* 6194 6985: contig of 792 bp in length
* 6986 7085: gap of 100 bp
* 7086 7897: contig of 812 bp in length
* 7898 7997: gap of 100 bp
* 7998 8867: contig of 770 bp in length
* 8868 9648: contig of 781 bp in length
* 9649 9748: gap of 100 bp
* 9749 10533: contig of 785 bp in length
* 10534 10633: gap of 100 bp
* 10634 11395: contig of 762 bp in length
* 11396 11495: gap of 100 bp
* 11496 12270: contig of 775 bp in length
* 12271 12370: gap of 100 bp
* 12371 13151: contig of 781 bp in length
* 13152 13251: gap of 100 bp
* 13252 14045: contig of 794 bp in length
* 14046 14145: gap of 100 bp
* 14146 14930: contig of 785 bp in length
* 14931 15030: gap of 100 bp
* 15031 15825: contig of 795 bp in length
* 15826 15925: gap of 100 bp
* 15926 16706: contig of 781 bp in length
* 16707 16806: gap of 100 bp
* 16807 17599: contig of 789 bp in length
* 17599 17695: gap of 100 bp
* 17696 18486: contig of 791 bp in length
* 18487 18586: gap of 100 bp
* 18587 19361: contig of 775 bp in length
* 19362 19461: gap of 100 bp
* 19462 20243: contig of 782 bp in length
* 20244 20343: gap of 100 bp
* 20344 21126: contig of 783 bp in length
* 21127 21226: gap of 100 bp
* 21227 22013: contig of 787 bp in length
* 22014 22113: gap of 100 bp
* 22114 22888: contig of 775 bp in length
* 22889 22988: gap of 100 bp
* 22989 23775: contig of 787 bp in length
* 23776 23875: gap of 100 bp
* 23876 24674: contig of 799 bp in length
* 24675 24774: gap of 100 bp
* 24775 25565: contig of 791 bp in length
```



* 57741	58524:	contig of 784 bp in length
* 58525	58624:	gap of 100 bp
* 58625	59404:	contig of 780 bp in length
* 59405	59504:	gap of 100 bp
* 59505	60291:	contig of 787 bp in length
* 60292	60391:	gap of 100 bp
* 60392	61181:	contig of 790 bp in length
* 61182	61281:	gap of 100 bp

Query Match            7.5%    Score 33.4; DB 73; Length 66534;  
Best Local Similarity    53.6%; Pred.No. 23;  
Matches    67; Conservative    0; Mismatches    58; Indels    0; Gaps

```
Qy   317 ctctcctgctgtactcgctggcagcatagtctgttgcctgctgctgtctgtcgttcacacctgt 376
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   23886 CTGCGTCTGTCGCACACTAAAGCATCCCNCTGCCAACCTGTCTTTGGNAGAAATGT 23945

Qy   377 gtactaacgtagtctgtccctcttgtagtggtagtactactctttagccatgtagctgtgc 436
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   23946 CTAATCAGGTTCAATTGCTCAATTTTAAATTTGGTGTGTTTTTGCCATCTGAGTTGTGTG 24005

Qy   437 tgttc 441
      ||||
Db   24006 AGTTC 24010
```

RESULT 34  
AC022781/c

LOCUS AC022781.5 GI:10305259  
DEFINITION Mus musculus chromosome 11 clone RP23-199H17 map 11, WORKING DRA  
SEQUENCE, 37 unordered pieces.  
ACCESSION AC022781  
VERSION AC022781.5  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mu

REFERENCE 1 (bases 1 to 200729)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Mus musculus chromosome 11, clone RP23-199H17  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 200729)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., An  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,  
Choepeil,T., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,  
DeAerlano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J.,  
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.  
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,  
Tirrell,A., Vassilieff,H., Viel,R., Vo,A., Wu,A., Wyman,D., Ye,W.,  
Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genom  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Sep 26, 2000 this sequence version replaced gi:9154653.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center -----  
Center: Whitehead Institute/ MIT Center for Genome Research  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information -----

```
Center project name: L5913
Center clone name: 199_H_17
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 180464 bases at least Q40
Consensus quality: 190837 bases at least Q30
Consensus quality: 194945 bases at least Q20
Insert size: 206000; agarose-fp
Insert size: 197129; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 7021: contig of 7021 bp in length
*
* 7022 7121: gap of 100 bp
* 7122 8310: contig of 1189 bp in length
*
* 8311 8410: gap of 100 bp
* 8411 9411: contig of 1001 bp in length
*
* 9412 9511: gap of 100 bp
* 9512 11581: contig of 2070 bp in length
*
* 11582 11681: gap of 100 bp
* 11682 13342: contig of 1661 bp in length
*
* 13343 13442: gap of 100 bp
* 13443 16024: contig of 2582 bp in length
*
* 16025 16124: gap of 100 bp
* 16125 17440: contig of 1316 bp in length
*
* 17441 17540: gap of 100 bp
* 17541 18922: contig of 1382 bp in length
*
* 18923 19022: gap of 100 bp
* 19023 21353: contig of 2331 bp in length
*
* 21354 21453: gap of 100 bp
* 21454 23674: contig of 2221 bp in length
*
* 23675 23774: gap of 100 bp
* 23775 25868: contig of 2094 bp in length
*
* 25869 25968: gap of 100 bp
* 25969 27966: contig of 1998 bp in length
*
* 27967 28066: gap of 100 bp
* 28067 31051: contig of 2985 bp in length
*
* 31052 31151: gap of 100 bp
* 31152 32336: contig of 2085 bp in length
*
* 32337 33336: gap of 100 bp
* 33337 36285: contig of 2949 bp in length
*
* 36286 36385: gap of 100 bp
* 36386 39476: contig of 3091 bp in length
*
* 39477 39576: gap of 100 bp
* 39577 42584: contig of 3008 bp in length
*
* 42585 42684: gap of 100 bp
* 42685 46164: contig of 3480 bp in length
*
* 46165 46264: gap of 100 bp
* 46265 49973: contig of 3709 bp in length
*
* 49974 50073: gap of 100 bp
* 50074 55389: contig of 5316 bp in length
*
* 55390 55489: gap of 100 bp
* 55490 59283: contig of 3794 bp in length
*
* 59284 59383: gap of 100 bp
* 59384 62877: contig of 3494 bp in length
*
* 62878 62977: gap of 100 bp
* 62978 68182: contig of 5205 bp in length
*
* 68183 68282: gap of 100 bp
* 68283 73862: contig of 5580 bp in length
*
* 73863 73962: gap of 100 bp
* 73963 83036: contig of 9076 bp in length
*
* 83039 83138: gap of 100 bp
* 83139 87822: contig of 4684 bp in length
*
*
* 87823 87922: gap of 100 bp
* 87923 92086: contig of 4164 bp in length
*
* 92087 92186: gap of 100 bp
* 92187 97007: contig of 4821 bp in length
*
* 97008 97107: gap of 100 bp
* 97108 123018: contig of 25911 bp in length
*
* 123019 123118: gap of 100 bp
* 123119 129307: contig of 6189 bp in length
*
* 129308 129407: gap of 100 bp
* 129408 137163: contig of 7756 bp in length
*
* 137164 137263: gap of 100 bp
* 137264 147510: contig of 10247 bp in length
*
* 147511 147610: gap of 100 bp
* 147611 155502: contig of 7892 bp in length
*
* 155503 155602: gap of 100 bp
* 155603 168939: contig of 13337 bp in length
*
* 168940 169039: gap of 100 bp
* 169040 181336: contig of 12297 bp in length
*
* 181337 181436: gap of 100 bp
* 181437 195395: contig of 13959 bp in length
*
* 195396 195495: gap of 100 bp
* 195496 200729: contig of 5234 bp in length.
*
FEATURES
  Location/Qualifiers
    source
      1..200729
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /chromosome="11"
        /map="11"
        /clone="RP23-199H17"
        /clone_lib="RPCI-23 Female Mouse BAC"
    misc_feature
      1..7021
        /note="assembly_fragment"
        clone_end:SP6
        vector_side:left
      7122..8310
        /note="assembly_fragment"
      8411..9411
        /note="assembly_fragment"
      9512..11581
        /note="assembly_fragment"
      11682..13342
        /note="assembly_fragment"
      13443..16024
        /note="assembly_fragment"
      16125..17440
        /note="assembly_fragment"
      17541..18922
        /note="assembly_fragment"
      19023..21353
        /note="assembly_fragment"
      21454..23674
        /note="assembly_fragment"
      23775..25868
        /note="assembly_fragment"
      25969..27966
        /note="assembly_fragment"
      28067..31051
        /note="assembly_fragment"
      31152..33236
        /note="assembly_fragment"
      33337..36285
        /note="assembly_fragment"
      36386..39476
        /note="assembly_fragment"
      39577..42584
        /note="assembly_fragment"
      42685..46164
        /note="assembly_fragment"
      46265..49973
        /note="assembly_fragment"
      50074..55389
        /note="assembly_fragment"
      55490..59283
        /note="assembly_fragment"
      59384..62877
        /note="assembly_fragment"
      62878..68182
        /note="assembly_fragment"
      68183..73862
        /note="assembly_fragment"
      73863..83036
        /note="assembly_fragment"
      83039..83138
        /note="assembly_fragment"
      83139..87822
        /note="assembly_fragment"
```



misc\_feature /note="assembly\_fragment:00579  
fragment\_chain:1"  
16325..19960  
/note="assembly\_fragment:00196  
fragment\_chain:2"  
20061..24131  
/note="assembly\_fragment:00024  
fragment\_chain:2"  
24232..28443  
/note="assembly\_fragment:01340  
fragment\_chain:2"  
28544..31045  
/note="assembly\_fragment:00423  
fragment\_chain:3"  
31146..36931  
/note="assembly\_fragment:01164  
fragment\_chain:3"  
37032..42061  
/note="assembly\_fragment:00612  
fragment\_chain:3"  
42162..44625  
/note="assembly\_fragment:00259  
fragment\_chain:4"  
44726..47635  
/note="assembly\_fragment:01001  
fragment\_chain:4"  
47736..51054  
/note="assembly\_fragment:00296  
fragment\_chain:5"  
51155..57015  
/note="assembly\_fragment:00190  
fragment\_chain:5"  
57116..64974  
/note="assembly\_fragment:00875  
fragment\_chain:6"  
65075..71197  
/note="assembly\_fragment:00537  
fragment\_chain:6"  
71298..74488  
/note="assembly\_fragment:00887  
fragment\_chain:7"  
74589..79354  
/note="assembly\_fragment:00909  
fragment\_chain:7"  
79455..85107  
/note="assembly\_fragment:00022"  
85208..89243  
/note="assembly\_fragment:00118"  
89344..92074  
/note="assembly\_fragment:00276"  
92175..99269  
/note="assembly\_fragment:00380"  
99370..103711  
/note="assembly\_fragment:00586"  
103812..106764  
/note="assembly\_fragment:00685"  
106865..113063  
/note="assembly\_fragment:00693"  
113164..118481  
/note="assembly\_fragment:00739"  
118582..122075  
/note="assembly\_fragment:00905"  
122176..128219  
/note="assembly\_fragment:01254"  
BASE COUNT 38763 a 24884 c 24158 g 37809 t 2605 others  
ORIGIN

Query Match 7.5%; Score 33.2; DB 80; Length 128219;  
Best Local Similarity 67.1%; Pred. No. 26;  
Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;  
Qy 234 ttactgttagcgttgacttcctgtgtactccgttgagccatttgagtct 293

Db 91890 TGTCTCTTCTTCGTAGAGAGTCCCTTGCTGTGATTCCTTGATACATTTAGGACT 91831  
Qy 294 gttaccgtg 303  
Db 91830 TTCTTCGCTG 91821  
RESULT 36  
AL513283 180381 bp DNA HTG 07-FEB-2001  
LOCUS Homo sapiens chromosome 1 clone RP11-95P13, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 17 unordered pieces.  
AL513283  
ACCESSION AL513283.3 GI:12718173  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 180381)  
AUTHORS Pavitt, R.  
TITLE Direct Submission  
JOURNAL Submitted (06-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Feb 8, 2001 this sequence version replaced gi:12619151.  
COMMENT ----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BA95P13  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 172799 bases at least Q40  
Consensus quality: 175859 bases at least Q30  
Consensus quality: 177452 bases at least Q20  
Insert size: 178781; sum-of-contigs  
Insert size: 175883; 17.3% error; agarose-fp  
Quality coverage: 4.04x in Q20 bases; sum-of-contigs Quality  
coverage: 4.32x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 11197: contig of 11197 bp in length  
\* 11198 11297: gap of 100 bp  
\* 11298 54629: contig of 43332 bp in length  
\* 54630 54729: gap of 100 bp  
\* 54730 65732: contig of 11003 bp in length  
\* 65733 65832: gap of 100 bp  
\* 65833 69227: contig of 3395 bp in length  
\* 69228 69327: gap of 100 bp  
\* 69328 72585: contig of 3258 bp in length  
\* 72586 72685: gap of 100 bp  
\* 72686 77520: contig of 4835 bp in length  
\* 77521 77620: gap of 100 bp  
\* 77621 87310: contig of 9690 bp in length  
\* 87311 87410: gap of 100 bp  
\* 87411 90165: contig of 2755 bp in length  
\* 90166 90265: gap of 100 bp  
\* 90266 130408: contig of 40143 bp in length  
\* 130409 130508: gap of 100 bp  
\* 130509 132537: contig of 2029 bp in length

```

* 132538 132637: gap of 100 bp
* 132638 138569: contig of 5932 bp in length
* 138570 138669: gap of 100 bp
* 138670 141380: contig of 2711 bp in length
* 141381 141480: gap of 100 bp
* 141481 151559: contig of 10079 bp in length
* 151560 151659: gap of 100 bp
* 151660 154357: contig of 2698 bp in length
* 154358 154457: gap of 100 bp
* 154458 162154: contig of 7697 bp in length
* 162155 162254: gap of 100 bp
* 162255 170186: contig of 7932 bp in length
* 170187 170286: gap of 100 bp
* 170287 180381: contig of 10095 bp in length.
FEATURES
    source
        1. .180381
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-95P13"
            /clone_lib="RPC1-11.1"
        1. .11197
            /note="assembly_fragment:00204"
            fragment_chain:1
            clone_end:SP6
            vector_side:left
            11298..54629
            /note="assembly_fragment:00098"
            fragment_chain:1
            54730..65732
            /note="assembly_fragment:00839"
            fragment_chain:2
            65833..69227
            /note="assembly_fragment:01812"
            fragment_chain:2
            69328..72595
            /note="assembly_fragment:00711"
            fragment_chain:2
            72686..77520
            /note="assembly_fragment:01336"
            fragment_chain:2
            77621..87310
            /note="assembly_fragment:01246"
            fragment_chain:2
            87411..90165
            /note="assembly_fragment:01290"
            fragment_chain:2
            90266..130408
            /note="assembly_fragment:00070"
            fragment_chain:2
            130509..132537
            /note="assembly_fragment:01353"
            fragment_chain:2
            132638..138569
            /note="assembly_fragment:00925"
            fragment_chain:3
            138670..141380
            /note="assembly_fragment:01044"
            fragment_chain:3
            141481..151559
            /note="assembly_fragment:01420"
            fragment_chain:3
            151660..154357
            /note="assembly_fragment:01190"
            fragment_chain:4
            154458..162154
            /note="assembly_fragment:01090"
            fragment_chain:4
            162255..170186
            /note="assembly_fragment:00648"
            170287..180381
            /note="assembly_fragment:01673"
BASE COUNT 53764 a 34451 c 34528 g 56033 t 1605 others

```

## ORIGIN

```

Query Match          7.5%; Score 33.2; DB 81; Length 180381;
Best Local Similarity 67.1%; Pred. No. 27;
Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 234 ttactgtgtgagcgttgatgactctctgctgctactccgcttgagccatttgagtgct 293
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 142733 TGTCCTCTTCTTCGCTAGAGGAGTCCTCGTGTGATTCCTCTTGATACATTTAGGACT 142792

QY 294 gttaccgctg 303
      |||||
Db 142793 TTCTCCGCTG 142802

RESULT 37
AC023785/209599 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 6 clone RP11-507C10, WORKING DRAFT
DEFINITION SEQUENCE, 18 unordered pieces.
ACCESSION AC023785
VERSION AC023785.3 GI:8099307
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Waterston, R.H.
            1 (bases 1 to 209599)
REFERENCE The sequence of Homo sapiens clone
            Unpublished
            2 (bases 1 to 209599)
            Waterston, R.H.
            Direct Submission
            Submitted (17-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On May 27, 2000 this sequence version replaced gi:7109627.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: HNH0507C10
----- Summary Statistics -----
Sequencing vector: M13; 78%
Sequencing vector: plasmid; 22%
Chemistry: Dye-primer ET; 78% of reads
Chemistry: Dye-terminator Big Dye; 22% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 199546 bases at least Q40
Consensus quality: 202608 bases at least Q30
Consensus quality: 204160 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 207899; sum-of-contigs
Quality coverage: 6.01 in Q20 bases; agarose-fp
Quality coverage: 6.11 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1478: contig of 1478 bp in length
* 1479 1578: gap of unknown length
* 1579 3500: contig of 1922 bp in length
* 3501 3600: gap of unknown length
* 3601 6635: contig of 3035 bp in length

```



```

/note="assembly_fragment:00039
fragment_chain:1"
misc_feature
157607..182957
/note="assembly_fragment:00361
fragment_chain:1"
misc_feature
183058..185348
/note="assembly_fragment:02093
fragment_chain:1"
misc_feature
185449..192902
/note="assembly_fragment:00794
fragment_chain:1"
misc_feature
193003..217060
/note="assembly_fragment:01647
fragment_chain:1"

BASE COUNT 57477 a 50328 c 49161 g 59193 t 901 others
ORIGIN

Query Match 7.4%; Score 32.8; DB 82; Length 217060;
Best Local Similarity 54.0%; Pred. No. 35;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 43 atctaagaaagcagtcctgccttctgggtcgctcctcaaacatgcctcctcgagcgc 102
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 195184 ACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 195125
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 103 ataataagaaagcagtcctgccttctgggtcgctcctcaaacatgcctcctcgagcgc 162
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 195124 ATCTTGAAAAAACTGTAACACTGTCATTCGCCCTGCACACGCTCTCATCTGGCTTGACAGG 195065
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 163 ggtg 166
|||
Db 195064 GGCG 195061

RESULT 39
AC023681
LOCUS AC023681 157890 bp DNA HTG 29-MAY-2000
DEFINITION Drosophila melanogaster clone RPCI98-39P3, WORKING DRAFT SEQUENCE,
8 unordered pieces.
ACCESSION AC023681
VERSION AC023681.3 GI:8101328
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 157890)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Taney,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,M., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejowski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Suegang,R.,
Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Waibah,M.,
Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 157890)
AUTHORS Worley,K.C.
TITLE Direct Submission

Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 29, 2000 this sequence version replaced gi:6997251.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: DHR
Center clone name: RPCI98-39P3
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 148728 bases at least Q40
Consensus quality: 154568 bases at least Q30
Estimated insert size: 133000; agarose-fp estimation
Estimated insert size: 157182; sum-of-contigs estimation
Quality coverage: 5.3x in Q20 bases; agarose-fp estimation
Quality coverage: 4.5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 47574: contig of 47574 bp in length
* 47575 47674: gap of unknown length
* 47675 79405: contig of 31731 bp in length
* 79406 79505: gap of unknown length
* 79506 107003: contig of 27498 bp in length
* 107004 107104: gap of unknown length
* 107104 133489: contig of 26386 bp in length
* 133490 133589: gap of unknown length
* 133590 153613: contig of 20024 bp in length
* 153614 153713: gap of unknown length
* 153714 155467: contig of 1754 bp in length
* 155468 155567: gap of unknown length
* 155568 156667: contig of 1100 bp in length
* 156668 156767: gap of unknown length
* 156768 157890: contig of 1123 bp in length.
FEATURES
Location/Qualifiers
1..157890
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RPCI98-39P3"
60176..60275
/note="assembly_fragment:contig16;
vector:Vector_pBACE3.6;
vector_side:left"
misc_feature
/note="assembly_fragment:contig17;
complement(87790..87889)
vector:Vector_pBACE3.6;
vector_side:right"
misc_feature
BASE COUNT 45713 a 32660 c 32796 g 46003 t 718 others
ORIGIN

Query Match 7.3%; Score 32.6; DB 68; Length 157890;
Best Local Similarity 58.9%; Pred. No. 41;
Matches 56; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 209 tgttatcgagccactgctgtctgttactgtgacgctgactccttgcgtga 268
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10445 TGTGTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 10504
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 269 ctccgcttgagccatttgagtgctgtttaccgtg 303

```





\* consists of 41 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 22725: contig of 22725 bp in length  
\* 22726 22825: gap of 100 bp  
\* 22826 27666: contig of 4841 bp in length  
\* 27667 27766: gap of 100 bp  
\* 27767 34399: contig of 6633 bp in length  
\* 34400 34499: gap of 100 bp  
\* 34500 37560: contig of 3061 bp in length  
\* 37561 37660: gap of 100 bp  
\* 37661 65584: contig of 27924 bp in length  
\* 65585 65684: gap of 100 bp  
\* 65685 69262: contig of 3578 bp in length  
\* 69263 69362: gap of 100 bp  
\* 69363 73709: contig of 4347 bp in length  
\* 73710 73809: gap of 100 bp  
\* 73810 83068: contig of 9259 bp in length  
\* 83069 83168: gap of 100 bp  
\* 83169 85224: contig of 2056 bp in length  
\* 85225 85324: gap of 100 bp  
\* 85325 87637: contig of 2313 bp in length  
\* 87638 87737: gap of 100 bp  
\* 87738 93159: contig of 5422 bp in length  
\* 93160 93259: gap of 100 bp  
\* 93260 98655: contig of 5396 bp in length  
\* 98656 98755: gap of 100 bp  
\* 98756 105031: contig of 6276 bp in length  
\* 105032 105131: gap of 100 bp  
\* 105132 111758: contig of 6627 bp in length  
\* 111759 111858: gap of 100 bp  
\* 111859 120085: contig of 8227 bp in length  
\* 120086 120185: gap of 100 bp  
\* 120186 123521: contig of 3336 bp in length  
\* 123522 123621: gap of 100 bp  
\* 123622 135166: contig of 11545 bp in length  
\* 135167 135266: gap of 100 bp  
\* 135267 139159: contig of 3893 bp in length  
\* 139160 139259: gap of 100 bp  
\* 139260 143355: contig of 4096 bp in length  
\* 143356 143455: gap of 100 bp  
\* 143456 153317: contig of 9862 bp in length  
\* 153318 153417: gap of 100 bp  
\* 153418 156147: contig of 2730 bp in length  
\* 156148 156247: gap of 100 bp  
\* 156248 159234: contig of 2987 bp in length  
\* 159235 159334: gap of 100 bp  
\* 159335 163807: contig of 4473 bp in length  
\* 163808 163907: gap of 100 bp  
\* 163908 167369: contig of 3462 bp in length  
\* 167370 167469: gap of 100 bp  
\* 167470 172739: contig of 5270 bp in length  
\* 172740 172839: gap of 100 bp  
\* 172840 186162: contig of 13323 bp in length  
\* 186163 186262: gap of 100 bp  
\* 186263 193526: contig of 7264 bp in length  
\* 193527 193626: gap of 100 bp  
\* 193627 195981: contig of 2355 bp in length  
\* 195982 196081: gap of 100 bp  
\* 196082 202543: contig of 6462 bp in length  
\* 202544 202643: gap of 100 bp  
\* 202644 216446: contig of 13803 bp in length  
\* 216447 216546: gap of 100 bp  
\* 216547 231463: contig of 14917 bp in length  
\* 231464 231563: gap of 100 bp  
\* 231564 234196: contig of 2633 bp in length  
\* 234197 234296: gap of 100 bp  
\* 234297 239413: contig of 5117 bp in length  
\* 239414 239513: gap of 100 bp

\* 239514 242362: contig of 2849 bp in length  
\* 242363 242462: gap of 100 bp  
\* 242463 255870: contig of 13408 bp in length  
\* 255871 255970: gap of 100 bp  
\* 255971 259826: contig of 3856 bp in length  
\* 259827 259926: gap of 100 bp  
\* 259927 276107: contig of 16181 bp in length  
\* 276108 276207: gap of 100 bp  
\* 276208 293707: contig of 17500 bp in length  
\* 293708 293807: gap of 100 bp  
\* 293808 301291: contig of 7484 bp in length  
\* 301292 301391: gap of 100 bp  
\* 301392 311585: contig of 10194 bp in length  
\* 311586 311685: gap of 100 bp  
\* 311686 313750: contig of 2065 bp in length.  
FEATURES  
source  
1. .313750  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-265C7"  
/clone\_lib="RPC1-11.1"  
1. .22725  
/note="assembly\_fragment:00276  
fragment\_chain:1"  
22826.27666  
/note="assembly\_fragment:00733  
fragment\_chain:1"  
27767.34399  
/note="assembly\_fragment:00703  
fragment\_chain:1"  
34500.37560  
/note="assembly\_fragment:00299  
fragment\_chain:1"  
37661.65584  
/note="assembly\_fragment:00236  
fragment\_chain:1"  
65685.69262  
/note="assembly\_fragment:01460  
fragment\_chain:1"  
69363.73709  
/note="assembly\_fragment:03799  
fragment\_chain:2"  
73810.83068  
/note="assembly\_fragment:00032  
fragment\_chain:2"  
83169.85224  
/note="assembly\_fragment:01349  
fragment\_chain:2"  
85325.87637  
/note="assembly\_fragment:01736  
fragment\_chain:2"  
87738.93159  
/note="assembly\_fragment:01701  
fragment\_chain:3"  
93260.98655  
/note="assembly\_fragment:01843  
fragment\_chain:3"  
98756.105031  
/note="assembly\_fragment:02264  
fragment\_chain:3"  
105132.111758  
/note="assembly\_fragment:00718  
fragment\_chain:4"  
111859.120085  
/note="assembly\_fragment:02459  
fragment\_chain:4"  
120186.123521  
/note="assembly\_fragment:01408  
fragment\_chain:5"  
123622.135166  
/note="assembly\_fragment:03466  
fragment\_chain:5"

```

misc_feature 135267..139159
/note="assembly_fragment:01967
fragment_chain:6"
misc_feature 139260..143355
/note="assembly_fragment:03108
fragment_chain:6"
misc_feature 143456..153317
/note="assembly_fragment:02873
fragment_chain:7"
misc_feature 153418..156147
/note="assembly_fragment:02323
fragment_chain:7"
misc_feature 156248..159234
/note="assembly_fragment:02935
fragment_chain:8"
misc_feature 159335..163807
/note="assembly_fragment:02794

Query Match 7.3%; Score 32.6; DB 79; Length 313750;
Best Local Similarity 52.6%; Pred. No. 41;
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 235 ttactgtgtacggttgatgactctctctgtgtgtactcgtgttgagccatttgagtgctg 294
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229989 TTTCTTTTGTCTCTCTGGCTCTGCCCTCTCTCCCTCTCCCTCGCCCGCTGAGCGCTG 229930

Qy 295 ttaccgtgagcactgaagactctctgtctgtactcgtggtgagcagcagcagcagcagc 354
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229929 TCTGACGGGACCCGACGACTCTCCCTGCTCTCCCTGGCAGCTGACGACTCTCTCTGTT 229870

Qy 355 tgcctggtctgggga 369
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229869 CTCCTGACTGCAGA 229855

RESULT 42
CLU20353/c 1983 bp mRNA VRT 31-DEC-1995
LOCUS
DEFINITION Columba livia growth hormone receptor (ghr) mRNA, complete cds.
ACCESSION U20353
VERSION U20353.1 GI:1142643
KEYWORDS
SOURCE domestic pigeon.
ORGANISM Columba livia
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Columbiiformes; Columbidae; Columba.
1 (bases 1 to 1983)
Cheng, C.H.K., Shaw, P.C., Tsim, K.W.K. and Lau, K.F.
Molecular cloning of the entire coding sequence of pigeon growth
hormone receptor cDNA by PCR techniques
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1983)
AUTHORS Cheng, C.H.K.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-1995) Christopher H.K. Cheng, Biochemistry, The
Chinese University of Hong Kong, Shatin, New Territories, Hong Kong
FEATURES
source
1. .1983
/organism="Columba livia"
/db_xref="taxon:9932"
/clone="pghr"
/sex="female"
/tissue_type="liver"
/dev_stage="adult"
1. .46
47. .1982
/genes="ghr"
/notes="encodes extracellular domain"
47. .1982
/genes="ghr"
/codon_start=1

5'UTR
gene
misc_feature
CDS

```

```

/function="transduction of growth hormone action"
/evidence="experimental"
/product="growth hormone receptor"
/protein_id="AA84745.1"
/db_xref="GI:1142644"
/translation="MDLRHLLLVLCANDSLASDDVLRLPQIKCRSELETFSC
YWTGDFYNLSAGFTIQLLYMKRNDWKECPTIYAGENSCTFNTSYTIWIYCVK
LVNKDEVDFDEKFSVDEIVLPDPVHLNMTLLNTSQTGIHDIQVRWDPPTADVQKG
WITLEYLOYKEVNETKWELEPLTMVPLYSLKIGRDEIRVRSQRTSEKMFGEFS
EILVVSQAGIEFVHCATEFPFWLWIFGACGLAVTILLLSKQSRKMLIFPP
VPVKIKGIDPDLLKGLDEVASILASHDNKYKLTOLVNDLLWVEFIELDIEDPEKNR
VSDTRLLSEHLKSHCLGAKDDSGRSCCEPDIPETDPSADTCDALSDIDQFKK
VTEKEDLLCLGRKNDDESPLANTDTQPRMSSTRPENQPPFPFADSDAASPSAH
NQLSNQSLRNTDFYAGVSDITPAGSVLSPGQSKVARARCEFTQNTFLDNAYFC
EADVKCIAVISHHEEDPRVQAQCNEDTYFTTESLTITGTSIGASTAETPSPEVPVP
DYTSIHIVHSPQGLVLNATALPVPDKFENKSGVYSTDQLNKIMP"
misc_feature 767..838
/genes="ghr"
/notes="encodes transmembrane domain"
misc_feature 839..1879
/genes="ghr"
/notes="encodes intracellular domain"
3'UTR 1883..1983
BASE COUNT 596 a 419 c 447 g 521 t
ORIGIN

Query Match 7.3%; Score 32.4; DB 8; Length 1983;
Best Local Similarity 48.0%; Pred. No. 43;
Matches 123; Conservative 0; Mismatches 131; Indels 2; Gaps 1;

Qy 97 caggcacaatggaaacagtcctctgccccttttgagggtcgtctcaaacatgccctcga 156
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1787 CAAGCCCTGTGGAGAGTGGACGATATGAAAGAGCTGATGCCAGCTAGTGTAAAGCTTTCTG 1668

Qy 157 gacgtcgg--tgaagccgtagttgttacaagactaacaacgccaatgagcgtcgatgttat 214
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1727 GACTTGGGGTTTCTGCTGTTGAAGCTCCAGACTGATGCCAGCTAGTGTAAAGCTTTCTG 1668

Qy 215 cggagccattcgtgctgttactgtgttagcgttgtagcttctctctgtactcgcg 274
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1667 TGGTGAAGTAAGTGTCTCTGTTACAGATTTCGGCTGAACACACGTGGCTCATCTCTCGT 1608

Qy 275 ttgagccatttgagctgtgtttaccgtgagcactgaagactcctctgtgtactgcg 334
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1607 GGGAAATCACAGCAATACATTTTTCACATCTGCCTCACAGAGTAGGCATTTGCCAGGG 1548

Qy 335 tggcagcatagctgt 350
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1547 TGAAGTTTGTCTGT 1532

RESULT 43
FRX297967
LOCUS FRX297967 2263 bp DNA PLN 01-DEC-2000
DEFINITION Fragaria x ananassa gene encoding methionine sulfoxide reductase.
ACCESSION AJ297967
VERSION AJ297967.1 GI:11342532
KEYWORDS methionine sulfoxide reductase.
SOURCE Fragaria x ananassa.
ORGANISM Fragaria x ananassa
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Rosales; Rosaceae; Fragaria.
1 (bases 1 to 2263)
Pedraza-Lopez, A., Cardenas-Torres, J. and Rodriguez-Franco, A.
Characterization of a peptide methionine sulfoxide reductase
isolated from strawberry fruit receptacles
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2263)
AUTHORS Rodriguez-Franco, A.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1996) Rodriguez-Franco A., Universidad de

```

Cordoba. Facultad de Ciencias, Bioquimica y Biologia Molecular, Av.

S. Alberto Magno S/N, Cordoba, SPAIN, 14071

FEATURES

Source

Location/Qualifiers

1..2263

/organism="Fragaria x ananassa"

/variety="Chandler"

/db\_xref="taxon:3747"

join(1372..1686,1813..2073)

/codon\_start=1

/product="methionine sulfoxide reductase"

/protein\_id="CAC17011.1"

/db\_xref="GI:11342533"

/translation="MASSTNNPALDSDTPENFGHELAQFASGCFWGSLEIRFORVV

GVIKEVGYSGHVDHPNYLVCGTTHSEVVRVQFQVCPYSDLLSVFWSRHDP

TNCGGDVGTQYRSGLVYVNEEQDLAKKSKKQEFKDKRVVTEILPAKFRYRAE

EYHQYLEKGGNGKNQSAQKCNPIKCYG"

<1372..1686

/number=1

1687..1812

/number=1

1813..>2073

/number=2

630 a 508 c 530 g 595 t

BASE COUNT

ORIGIN

Query Match 7.3%; Score 32.4; DB 14; Length 2263;

Best Local Similarity 60.0%; Pred. No. 43; Indels 0; Gaps 0;

Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 206 cgatgtatcgagccattcgtgtcttactgttagcgttgacttgccttgcgt 265

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1142 CTACGATACGGAGTCAGCAACTATGCTCTTGTGCTGTTGGACATACATCA 1201

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 266 gtactcgttgagccattgagtcgtgt 295

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1202 ATATTCTCTTGAACACATTGAACCGCTGT 1231

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 44

AC018863/C

LOCUS

DEFINITION

AC018863 186053 bp DNA HTG 12-AUG-2000

Homo sapiens chromosome 2 clone RP11-20J1, WORKING DRAFT SEQUENCE,

6 unordered pieces.

AC018863

VERSION

AC018863.2 GI:9755480

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 186053)

Waterston, R.H.

AUTHORS

TITLE

THE SEQUENCE OF HOMO SAPIENS CLONE

REFERENCE

1 (bases 1 to 186053)

Waterston, R.H.

AUTHORS

TITLE

Direct Submission

Submitted (21-DEC-1999) Genome Sequencing Center, Washington

University School of Medicine, 444 Forest Park Parkway, St. Louis,

MO 63108, USA

COMMENT

On Aug 9, 2000 this sequence version replaced gi:6624119.

Chemistry: Dye-terminator Big Dye; 10% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 180709 bases at least Q40

Consensus quality: 182200 bases at least Q30

Consensus quality: 183266 bases at least Q20

Insert size: 179000; agarose-fp

Insert size: 185553; sum-of-contigs

Quality coverage: 5.10 in Q20 bases; agarose-fp

Quality coverage: 5.09 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 6 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 6671: contig of 6671 bp in length

\* 6672 6771: gap of unknown length

\* 6772 18019: contig of 11248 bp in length

\* 18020 18119: gap of unknown length

\* 18120 33989: contig of 15870 bp in length

\* 33990 34089: gap of unknown length

\* 34090 56036: contig of 21947 bp in length

\* 56037 56136: gap of unknown length

\* 56137 97345: contig of 41209 bp in length

\* 97346 97445: gap of unknown length

\* 97446 186053: contig of 88608 bp in length.

FEATURES

source

1..186053

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="2"

/clone="RP11-20J1"

1..6671

/note="assembly\_name:Contig11"

6772..18019

/note="assembly\_name:Contig12"

18120..33989

/note="assembly\_name:Contig13

clone\_end:SP6

vector\_side:right"

34090..56036

/note="assembly\_name:Contig14

clone\_end:T7

vector\_side:right"

56137..97345

/note="assembly\_name:Contig15"

97446..186053

/note="assembly\_name:Contig16"

58625 a 33523 c 33285 g 60118 t 502 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 57; Conservative

0; Mismatches 41; Indels

0; Gaps 0;

QY 218 agccattcgtgtcttactgttagcgttgactctctgtactcctgtg 277

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 78936 AACATGCCCCCTTCTTGGATTGTAGCCTTGGAGATTCTGCTTACTCCGCTG 78877

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 278 gagccattgagtgctgtttaccgtgcagcactcgaag 315

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 78876 CAGTTGTTTCTCAGTATTAAATTTTGCCTCTCCCTCAG 78839

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 45

AC013799/c

LOCUS

DEFINITION

Homo sapiens clone RP11-20J1, WORKING DRAFT SEQUENCE, 11 unordered

pieces.

AC013799 190650 bp DNA HTG

16-MAR-2000

Search completed: September 7, 2001, 17:04:05  
Job time: 7509 sec



**THIS PAGE BLANK (USPTO)**

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**